Statistical depth outside the location setting

VAN BEVER, GERMAIN

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Citation for pulished version (HARVARD):
VAN BEVER, GERMAIN 2018, *Statistical depth outside the location setting: the case of scatter, shape and concentration parameters*.

Faculté des Sciences & Solvay Brussels School of Economics and Management (SBS-EM)

Département de Mathématique & European Center for Advanced Research in Economics and Statistics (ECARES)

Statistical depth outside the location setting: the case of scatter, shape and concentration parameters

Germain VAN BEVER

Université libre de Bruxelles

Mémoire présenté dans le cadre du Concours 2018 de la Classe des Sciences
(Groupe I - Mathématiques: a)

Mars 2018

"A mathematician, like a painter or a poet, is a maker of patterns. If his patterns are more permanent than theirs, it is because they are made with ideas."

Godfrey Harold Hardy, in Hardy (1940).

"Geometry is the right foundation of all painting."

Albrecht Dürer, in Dürer (1525).

Contents

Introduction

1 The notion of centrality — Depth

1.1 The origins – Multivariate medians

The notion of center of an object, be it a set of observations, a physical object or a random variable X, is difficult to define. Whether definitions refer to *a point, pivot or axis around which anything rotates or revolves*[1], therefore being inseparable from the notion of symmetry itself, or to *the middle point, as the point or part that is equally distant from all points, sides, ends or surfaces of something*[1] there is no canonical way to define it. All definitions, however, agree on the importance of distance or geometry in the construction of such notion.

From a mathematical point of view, many such notions of center—of a random variable X having distribution P, say—were defined in the univariate setting and an abundant literature of so-called *univariate location measures* exists. The most canonical notion, of course, is the mean or expectation $\mathrm{E}[X] = \int x dP(x)$. However, due to the high sensitivity of this particular location functional, it is often advocated that, should the focus be put on broader and more robust applicability, the competing notion of median presents much more appeal. Indeed, it is a well-known fact that, although it suffices to have a single point contaminating a data set and going to infinity to force the mean to do the same, the median will require, by contrast, 50% of the data to be moved to infinity before it does as well.

The geometry of the quadratic distances that underlines the definition of the mean (that, alternatively, can be defined as the location minimising the functional $x \mapsto \mathrm{E}[(X-x)^2]$) makes it particularly amenable to generalisation in higher dimensions. This is the reason why theory based upon the multivariate Gaussian distribution has been dominating multivariate analysis for a long time. Indeed, the multivariate normal distribution is the sole distribution for which the sample mean is the maximum likelihood estimate of its location parameter (a fact that can be traced back to Gauss, 1809).

The median of a random variable X, denoted $\mathrm{Med}(X)$, is defined through the cumulative distribution as the point m_P such that $P\big[X \leq m_P\big] \geq 1/2$ and $P\big[X \geq m_P\big] \leq 1/2$. Accordingly, the median $\mathrm{Med}(X^{(n)})$ of a dataset $X^{(n)} = \{X_1, \ldots, X_n\} \subset \mathbb{R}$ will be defined by substituting the empirical distribution on $X^{(n)}$, $P^{(n)}$ say, to P.

The lack of natural ordering in $\mathbb{R}^k$ prevents a straightforward extension of the latter definitions, so that one may wonder what the appropriate analogues in two or more dimensions are. Regardless of the notion employed, there are, however, certain properties these notions should satisfy, the first of which being, as in the univariate case, robustness[2]. Another such condition is that, under symmetry, the (multivariate) median should coincide with the symmetry center. In the univariate case, the notion of symmetry presents no ambiguity (a random vector X is symmetric about μ if $X - \mu \overset{\mathcal{D}}{=} \mu - X$, where $\overset{\mathcal{D}}{=}$ denotes equality in distribution) and univariate location measures typically coincide under symmetry. This is not necessarily so in higher dimensions as symmetry can be generalised in many ways, see Section 1.3 for details.

[1] Harrap's Dictionnary of Contemporary English.

[2] Many tools for measuring robustness exist. A classical way to compare robustness of location measures is through their breakdown point (see Hodges (1967) for the univariate definition and Hampel (1971) more generally).

Extending the concept of median to the multivariate setup (or a similar approach that consists in ordering multivariate observations) has generated numerous publications over the past few years; see, for example, Barnett (1976) or Hettmansperger *et al.* (1992) for a review about how to order multivariate data, and Donoho & Gasko (1987) or Small (1990) for a comprehensive summary of existing multivariate analogues of the median at that time.

The naive[3] attempt to take the componentwise median (first used by Hayford (1902) for geographical considerations) showed that, despite some encouraging robustness properties, very poor performances were to be expected, particularly in the case of highly correlated marginals. To add on this drawback, this vector of medians is not equivariant under rotation or arbitrary affine transformation of the data (see, for example, Bickel, 1964; Barnett, 1976), so that the way of measuring the data will have a strong impact on the outcome of the procedure (which is of course to be avoided).

To improve on this definition, many authors independently considered the spatial median[4] as a natural generalisation of the univariate median in different situations, see, e.g. Gini & Galvani (1929), Scates (1933) and Haldane (1948).

Definition 1.1. Let X be a random vector having distribution P on $\mathbb{R}^k$. The *spatial median* of X is the location $\hat{\mu}_S(P) \in \mathbb{R}^k$ that minimises $\mathrm{E}_P\left[||X-\mu||\right]$, where $||.||$ denotes the standard Euclidian norm.

Note that the distribution P may be that of the empirical distribution $P^{(n)}$ of n i.i.d. data points $X_1, \ldots, X_n$ sharing the same distribution P. Locating the spatial median, in that case, amounts to finding the solution of

$$\hat{\mu}^{(n)} = \mathrm{arginf}_{\mu \in \mathbb{R}^k} \sum_{i=1}^{n} ||X_i - \mu||.$$

This problem, for which there now exist plenty of algorithmic solutions (see, for example, Vardi & Zhang, 2001), is actually far much older than the introduction of the multivariate median. Indeed, minimising a weighted sum of the Euclidian distances from m points in $\mathbb{R}^k$ was already known, in industrial applications, as the optimal location problem of Weber (1909). The problem actually goes back to Fermat in the seventeenth century (for $m = 3$ and equal weights) but was only generalised to its actual form by Simpson (1750) (see Kuhn, 1973). It is interesting to note that Kemperman (1987), following the same idea, discussed the median of a finite measure on an arbitrary Banach space and proved, under strict convexity of the underlying space and provided the distribution is not supported on a straight line, uniqueness of the resulting location functional[5].

Other celebrated instances of multivariate medians include the simplicial volume median from Oja (1983) and a median-functional obtained through a "peeling" approach (see Barnett

[3]Actually, the second-most naive approach, if one considers the poorly defined tentative extension through the cdf, defining "a" (non-unique) median as a location m such that $P[X \leq m] \geq 1/2$ and $P[X \geq m] \leq 1/2$, for the partial order $x \leq y \Leftrightarrow x_i \leq y_i$ for each marginal i.

[4]The denomination *mediancenter* is used in Gower (1974) and the first reference as a "spatial median" can be found in Brown (1983).

[5]Uniqueness in the Euclidian case of $\mathbb{R}^k$ was treated by Milasevic & Ducharme (1987). This is in strict contrast with the univariate case where uniqueness does not hold in general.

(1976), the comment by Plackett (1976) and Green (1981)). Based on univariate measures of outlyingness, Stahel (1981) and, independently, Donoho (1982) used projection pursuit ideas to generalise the univariate weighted location estimator of Mosteller & Tukey (1977) and defined a robust location estimator as a weighted average, with weights based on outlyingness measures. A final example is the projection median, which was studied in Tyler (1994).

A classical requirement for location estimators/functionals is the *affine-equivariance property.*

Property 1.2. Let X be a random vector on $\mathbb{R}^k$ having distribution $P \in \mathcal{P}$ and $\hat{\mu}(P)$ be a multivariate location functional. Then $\hat{\mu}(.) : \mathcal{P} \to \mathbb{R}^k$ is said to be *affine-equivariant* if $\hat{\mu}\big(P_{AX+b}\big) = A\hat{\mu}(P) + b$, where P_{AX+b} denotes the distribution of $AX + b$ for the $k \times k$ invertible matrix A and $b \in \mathbb{R}^k$.

As it turns out, the median from Definition 1.1 is not affine-equivariant for a non-orthogonal matrix A in general. This is the reason why Chakraborty & Chaudhuri (1996, 1998) and Chakraborty *et al.* (1998) proposed a data-driven transformation-retransformation technique turning the spatial median into an affine-equivariant location functional. A similar approach was adopted in Hettmansperger & Randles (2002), where the initial data is first standardised using Tyler's M-estimator of scatter (Tyler, 1987).

1.2 Existing notions – Depth functions

Many of the definitions introduced in the previous section share a common construction. Most of them indeed define a multivariate median (in $\mathbb{R}^k$) as a location optimising some criterion, that, in some sense, reflects the centrality of a point x with respect to the underlying distribution. This motivated the development of general ways to measure centrality via *depth functions.* Such mappings provide, in turn, new multivariate medians. They also—and contrary to the naive approach to multivariate location that looks only for the most central point—allow for (i) comparing relative centrality of two locations and, consequently, (ii) providing a center-outward ordering (that would, in turn, make possible the definition of multivariate quantiles, see Serfling, 2002).

More precisely, letting $\mathcal{P}$ denote the class of distributions over the Borel sets $B \in \mathcal{B}^k$ of $\mathbb{R}^k$, a depth function is a mapping $D(.,.) : \mathbb{R}^k \times \mathcal{P} \to \mathbb{R} : (x, P) \mapsto D(x, P)$[6] that, intuitively, associates with any location x a value reflecting its centrality with respect to distribution P. Several recent reviews on data depth include Liu *et al.* (2006) (and in particular the introductive chapter by Serfling, 2006), Cascos (2009), Romanazzi (2009), Mosler (2012) or the theoretical approach from Zuo & Serfling (2000).

The most celebrated of these mappings, halfspace depth, is now described in its population version.

The earliest notion of depth dates back to Tukey (1975) (see also Tukey, 1977). Initially introduced as a tool to picture the data, the halfspace depth became increasingly popular and

[6]Some rare depth functions will only be defined for a subset of $\mathcal{P}$. Also, the notation $D(x, P) := D_P(x)$ is often used to emphasise the fact thatthe depth function measure the centrality of x for the underlying, often fixed, distribution P.

was quickly widely used in many procedures, due to its numerous useful properties and its intuitive interpretation.

In the univariate case, the median (of some distribution with cdf F) is univocally characterised as the location maximising $D(x, F) = \min\big(F(x), 1 - F(x^-)\big)$, where $F(x^-)$ denotes the left-sided limit of F at x. Generalising this last quantity to the multivariate case, the halfspace depth of $x \in \mathbb{R}^k$ is defined as the "minimal" probability of any closed halfspace containing x[7].

Definition 1.3. Let $x \in \mathbb{R}^k$. Let X be a random vector on $\mathbb{R}^k$ with distribution $P \in \mathcal{P}$. The *halfspace depth* of x with respect to P is

$$HD_P(x) := \inf_{u \in \mathcal{S}^{k-1}} P[u'(X - x) \geq 0].$$

Germ of this definition, in the bivariate case and only interested in the associated multivariate median, can be traced back to Hotelling (1929). The halfspace depth is actually a special case of particular applications used in economic game theory called "index functions"; see Small (1987). Rousseeuw & Ruts (1999) cover many of the properties of halfspace depth.

Many other such functions have been introduced in the literature. A (tentatively exhaustive, in historical order of introduction) list include simplicial depth (Liu, 1987, 1988, 1990), majority depth Liu & Singh (1993), projection depth (Zuo & Serfling, 2000; Hu *et al.*, 2011), Mahalanobis depth (Liu, 1992), zonoid depth (Koshevoy & Mosler, 1997), simplicial volume depth and L^p depth (Zuo & Serfling, 2000), spatial depth (Vardi & Zhang, 2000), spatial rank depth (Gao, 2003), spherical depth (Elmore *et al.*, 2006), and lens depth (Liu & Modarres, 2011).

A few other depth functions were defined elsewhere in the literature but were not considered in the list above as they do not meet one of the following natural requirements: (i) Although locating the center of a data set is important, the notion of depth should be defined as generally as possible and should, in particular, be able to deal with continuous distributions P. (ii) The argument is actually valid the other way around as depth should not be only limited to the latter type of distributions. (iii) Finally, depth functions that, even under the strongest hypothesis of symmetry on the distribution (that is, in circumstances where the center can be unequivocally defined) may fail to assign maximal value to the symmetry center should not be considered.

The incriminated depth functions were the likelihood/probing depth from Fraiman *et al.* (1997) and Fraiman & Meloche (1999), the interpoint distance depth from Lok & Lee (2011) (see also Bartoszyński *et al.*, 1997) or the convex hull peeling depth and the proximity depth (also known as Delaunay depth) from Hugg *et al.* (2006).

1.3 Statistical depth functions – A paradigmatic approach

Each depth function from the previous section has its own advantages and drawbacks (depending, also, on the objectives at hand), so that one might find it difficult to know which depth

[7] Note that the bivariate halfspace depth of a point is equivalent to the sign test statistic of Hodges (1955).

function to use. To discriminate between the many depth definitions, Zuo & Serfling (2000) stated four desirable properties that depth functions should ideally satisfy. Without loss of generality, only non-negative and bounded functions are considered. The four properties are

(P1) *affine-invariance*: The depth of a point $x \in \mathbb{R}^k$ should not depend on the underlying coordinate system nor on the scales used;

(P2) *Maximality at center*: For a symmetric distribution, the depth function should attain its maximum value at the center of symmetry;

(P3) *Monotonicity relative to deepest point*: For a distribution possessing a unique deepest point, the depth of a point $x \in \mathbb{R}^k$ should be decreasing as x moves away along any ray from that point;

(P4) *Vanishing at infinity*: The depth of a point $x \in \mathbb{R}^k$ should converge to zero as $||x||$ approaches infinity.

The notion of symmetry used in Property (P2), although defined unambiguously in the univariate case, may differ from one concept to another. They include, in decreasing order of generality, *halfspace symmetry* (X is halfspace symmetric about μ if $P[H] \geq 1/2$ for any closed halfspace containing μ), *angular symmetry* (X is angularly symmetric about μ if $(X-\mu)/||X-\mu|| \overset{\mathcal{D}}{=} (\mu-X)/||X-\mu||$), *central symmetry* ($X$ is centrally symmetric about μ if $X-\mu \overset{\mathcal{D}}{=} \mu-X$), and *spherical symmetry* (X is spherically symmetric about μ if $(X-\mu) \overset{\mathcal{D}}{=} O(X-\mu)$ for any orthogonal matrix O).

Let $\mathcal{P}$ denote the set of all distributions on $\mathbb{R}^k$ and P_X the distribution of the random vector X. In view of the previous requirements, Zuo & Serfling (2000) adopted the following definition of *statistical depth function*.

Definition 1.4. The bounded mapping $D(.,.) : \mathbb{R}^k \times \mathcal{P} \to \mathbb{R}^+$ is called a *statistical depth function* if it satisfies the four following properties :

(P1) for any $k \times k$ invertible matrix A, any k-vector b, any $x \in \mathbb{R}^k$ and any random vector $X \in \mathbb{R}^k$, $D(Ax+b, P_{AX+b}) = D(x, P_X)$;
(P2) if μ is a center of (central, angular or halfspace) symmetry of $P \in \mathcal{P}$, then it holds that $D(\mu, P) = \sup_{x \in \mathbb{R}^k} D(x, P)$;
(P3) for any $P \in \mathcal{P}$ having deepest point μ, $D(x, P) \leq D((1-\lambda)\mu + \lambda x, P)$ for any x in $\mathbb{R}^k$ and any $\lambda \in [0,1]$;
(P4) for any P, $D(x, P) \to 0$ as $||x|| \to \infty$.

Other proposals of such paradigmatic approach to depth functions have been introduced elsewhere in the literature. Comparison of depth functions based on different criterions, among which the "stochastic order preservation" was provided in Zuo (2003a). Also, Dyckerhoff (2002) (see also Mosler, 2012) did not use property (P2) but also added the technical property

(P5) *upper semicontinuity*: For any $P \in \mathcal{P}$, the upper level sets $D_\alpha(P) = \{x \in \mathbb{R}^k | D(x, P) \geq \alpha\}$ are closed for all $\alpha > 0$.

Under (P3), the sets $D_\alpha(P)$ (commonly known as the *depth regions*) are nested and star-shaped about the deepest point, should it exist. They are of particular interest, as they bring much information about the spread, shape and symmetry of the underlying distribution (see Serfling, 2004a). The *depth contours* (boundary of the depth regions) even characterise, under very mild conditions, the underlying distribution[8] (see Kong & Zuo (2010) and references therein). When the depth function $D(.,.)$ satisfies the more stringent assumption

(P3') *Quasiconcavity*: For any $P \in \mathcal{P}$, $D(., P)$ is a quasiconcave function, that is, its upper level sets $D_\alpha(P)$ are convex for all $\alpha > 0$,

the resulting depth function is often called a *convex statistical depth function.* Definition 1.4 plays a central role in the literature and is the most widely accepted abstract definition of statistical depth function. All depth functions introduced in the previous section are statistical depth functions in that sense, although some restrictions (on $\mathcal{P}$, on the symmetry used, etc.) may be required.

Upper semicontinuity (P5) was proved to actually hold for all the depth functions introduced here (see Liu & Singh, 1993; Mizera & Volauf, 2002; Mosler, 2012; Gao, 2003; Elmore *et al.*, 2006; Liu & Modarres, 2011). Results about convexity are more sparse. (P3') holds for halfspace depth (Rousseeuw & Ruts, 1999) but does not, for example, for simplicial depth.

2 Depths outside the location setting: regression and parametric depths

The successful story of depth in the location setup has motivated extending the concept to other parametric setups. Several proposals exist in the regression model and a full parametric approach to the notion of centrality has been developed in the early noughties.

2.1 Regression depths

Parallel to the extension from the univariate to the multivariate median, where a structural property of the one-dimensional median serves as ground to define a multidimensional counterpart, alternative characterizations of (halfspace) depth are required for proper generalisation. A first equivalence result can be found in Carrizosa (1996), where it is proved that

$$HD_P(x) = \inf_{y \in \mathbb{R}^k} P\Big(\big\{a : |y - a| \geq |x - a|\big\}\Big),$$

that is, the halfspace depth of x is the smallest probability (among all fixed choices of y) of the set of points that are closer to x than to y.[9] The latter equality allows extension to problems with non-Euclidian metrics or dissimilarity measures $\delta(x, y)$ by defining the depth of an element x as

$$HD_P(x) = \inf_y P\big[\{a : \delta(y, a) \geq \delta(x, a)\}\big].$$

[8]Partial results on the question whether the depth function uniquely determines the underlying distribution are available in the literature: see Struyf & Rousseeuw (1999); Koshevoy (2002, 2003); Mosler & Hoberg (2006); Hassairi & Regaieg (2008).

[9]This quantity is often used in the Operational Research literature to address facility location problem, much in the spirit of Weber's problem, see Section 1.1.

Parallel extension to the (single-output) regression setup goes as follows. Given a probability measure P on $\mathbb{R}^k \times \mathbb{R}$, corresponding to a multivariate random variable $(X,Y)'$, the depth of the hyperplane $H_{a,b} \equiv y = a'x + b$ is defined as

$$RD_P(H_{a,b}) = \inf_{(c,d)\in\mathbb{R}^k\times\mathbb{R}} P\Big(\big\{(x,y) \in \mathbb{R}^k \times \mathbb{R} : |y - a'x - b| \geq |y - c'x - d|\big\}\Big).$$

Both depths above received very little attention in the literature as only few properties were explored.

Equivalently, in the sample case, Tukey depth of x can also be seen as the minimal relative number of points that need to be removed before x ceases to be a Pareto optimum (for the distance function $f(x,.) = ||x - .||$) with respect to the remaining dataset[10]. This motivates the celebrated regression depth introduced in Rousseeuw & Hubert (1999), admitting the following definition in the sample case[11].

Definition 2.1. The *regression depth* $RD(H_{a,b}, P^{(n)})$ of an hyperplane $H_{a,b}$ with respect to the dataset $\{(x_i, y_i), 1 \leq i \leq n\} \subset \mathbb{R}^k \times \mathbb{R}$ (whose empirical distribution is denoted $P^{(n)}$) is the minimal relative number of points that need to be removed to make (a,b) a non-Pareto optimum of the residual function $f : (\mathbb{R}^k \times \mathbb{R})^2 \to \mathbb{R}^+ : \big((a,b),(\boldsymbol{x},y)\big) \to |y - a'x - b|$ with respect to the remaining dataset.

As usual, maximizing the depth function will provide a median-type estimate of regression. Bounds on the minimal depth of this estimator are provided, for the bivariate case, in Rousseeuw & Hubert (1999), through the construction of the "catline", an hyperplane with minimal depth 1/3. Although the population version of the latter definition does not seem easy to define, it will be given as a by-product of the general tangent depth introduced below.

2.2 Parametric depths

Mizera (2002) based on the same ideas of Pareto optimality a concept of *global depth*, that extends location and regression depths to an arbitrary parametric model. Consider a random k-vector X with a distribution $P = P_{\vartheta_0}$ in the parametric family $\mathcal{P} = \big\{P_\vartheta |\ \vartheta \in \Theta \subset \mathbb{R}^k\big\}$ (k may differ from k). Let $(\vartheta, x) \mapsto F_\vartheta(x)$ be a mapping that measures the "quality" of the parameter value θ for the observation x. A natural definition of depth is

Definition 2.2. Let $\vartheta \in \Theta$. The *global depth* of ϑ, $GD_{P^{(n)}}(\vartheta)$, with respect to the empirical distribution of the i.i.d random sample $X_1, \ldots, X_n$ with common distribution $P \in \mathcal{P}$ is the minimal relative number of points that need to be removed before ϑ is no longer a Pareto optimum of $f(\vartheta, x) = F_\vartheta(x)$ with respect to the remaining dataset.

Now, while this definition still remains uninspiringly limited to the sample case, the following restriction will allow a full treatment of parametric depth. Assuming that the objective function $F_\vartheta(x)$ is differentiable and convex with respect to ϑ, it is easy to show (see Mizera (2002) for details) that $GD_{P^{(n)}}(\vartheta) = \min_{||u||=1} \sharp\{i : u'\nabla_\vartheta F_\vartheta(X_i) \geq 0\} = HD_{P^{(n)}_{\nabla_\vartheta F}}(0_k)$, where

[10]Recall that a point x is a Pareto optimum for the function $f(\cdot,\cdot)$ with respect to some dataset A if there exists no y such that $f(y,a) \leq f(x,a)$ for all $a \in A$, with a strict inequality for at least one element of A.

[11]Population version is easy to obtain but lacks intuitive appeal hence is omitted here.

$P^{(n)}_{\nabla_\vartheta F}$ denotes the empirical distribution of $\nabla_\vartheta F_\vartheta(X_i)$, $1 \leq i \leq n$ and $0_k = (0, \ldots, 0)' \in \mathbb{R}^k$. This amounts to looking at the depth of 0_k among the directions $\nabla_\vartheta F_\vartheta(X_i)$, $i = 1, \ldots, n$, of maximal increase of $\vartheta \mapsto F_\vartheta(x)$.

The following concept then typically attributes large depth to "good" parameter values, that is, to parameter values ϑ that are close to ϑ_0.

Definition 2.3. The *tangent depth* of ϑ with respect to $P \in \mathcal{P}$ is $TD_P(\vartheta) = HD_{P_{\nabla_\vartheta F_\vartheta(\mathbf{X})}}(0)$, where $P_{\nabla_\vartheta F_\vartheta(X)}$ denotes the distribution of $\nabla_\vartheta F_\vartheta(X)$ under $X \sim P$.

Tangent depth reduces to the particular cases of classical (location) halfspace depth for $\vartheta = x$ and $f(x, y) = ||x - y||$, and of regression depth, for which $\vartheta = (a, b)$ and $f((a, b), (x, y))$ is as in Definition 2.1. Actually, any modification of the objective function F to $h(F)$, with $h : \mathbb{R}^+ \to \mathbb{R}^+$ smooth and monotone increasing would lead to the same definition.

In order to make explicit the link existing between halfspace depth and parametric depth in the location setting, we will adopt in the sequel the notation $HD_P(\theta)$, thereby seeing halfspace depth as a measure of appropriateness of the *location parameter* θ in P. Note the difference with ϑ, symbol which will be used in a generic parametric model.

Choosing appropriately the objective function $(\vartheta, x) \mapsto F_\vartheta(x)$ may be difficult in some setups. Denoting $L_\vartheta(x)$ for the likelihood function, the general, *likelihood-based*, approach consists in taking $F_\vartheta(x) = -\log L_\vartheta(x)$; see, e.g., Mizera & Müller (2004); Müller (2005). In the location and regression cases considered above, it can be seen that Gaussian or t_ν-likelihoods lead to the halfspace and regression depth, respectively.

A particular example of application of tangent depth can be found in Mizera & Müller (2004), where *location-scale* depth of $\vartheta = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+_0$ with respect to the univariate distribution P is developed, based on Definition 2.3, and explored. Interestingly, the proposed depth can be seen as the bivariate halfspace depth of projected observations in the Poincaré plane model embedded with the Lobachevski geometry.

3 Scatter, shape and concentration parameters

In the multivariate setting, the family of elliptical distributions plays a central role and has dominated statistical inference (in parametric models) for the most part of the last four decades. Recall that X is said to follow an elliptical distribution $P = P^X$ with location $\theta(\in \mathbb{R}^k)$ and scatter $\Sigma(\in \mathcal{P}_k$, the set of $k \times k$, positive definite matrices) if and only if $X \overset{\mathcal{D}}{=} \theta + \Sigma^{1/2} Z$, where $Z = (Z_1, \ldots, Z_k)'$ is spherically symmetric about the origin of $\mathbb{R}^k$ (that is, $OZ \overset{\mathcal{D}}{=} Z$ for any $k \times k$ orthogonal matrix O). To achieve identifiability of the model, an extra condition has to be imposed on the density of Z_1 and takes different form in the literature, depending on the statistical problem at hand, see Chapter I for one such condition.

The scatter parameter Σ entirely encodes the dispersion pattern of the distribution as well as the geometry of the density contours. Next to location parameters, scatters (as well as shape and concentration, see below) matrices are without a doubt the most central quantities in modern statistics. Understanding these parameters is crucial in many inference problems including, for example, linear and quadratic discriminant analysis, principal component analysis (PCA), etc. Note also that scatter functionals allow to define univocally scatter matrices in non-elliptical settings (Independent component analysis, see Oja *et al.* (2006) and references therein, being one instance).

It is therefore surprising that little efforts were made to provide depth notions for dispersion parameters. The necessity to compare scatters matrices was already acknowledged in Serfling (2004b). The author indeed calls for an extension of the Mizera & Müller (2004) location-scale depth concept described above into a location-scatter one in the multivariate setting.

The only depth in that direction is the one proposed by Zhang (2002) or the concept, close in spirit, from Chen *et al.* (2017). Both concepts, however, are not studied from a geometric point of view and, in particular, pay no attention to the depth regions and their properties. Doing so, most properties of the function are not explored.

Parallely, a scaled version of the scatter Σ, i.e. the associated *shape*, is the parameter of interest in many multivariate statistics problems. It is indeed often sufficient to know Σ up to a positive scalar factor $S(\Sigma)$, its *scale*, to conduct inference. These include, for example, PCA, for which the eigendirections are not affected by any scaling of the scatter matrix. Moreover, under the situation where second-order moments are infinite, shape – unlike Σ – remains well-defined and still fixes the geometry that permits PCA. To the best of our knowledge, there does not exist any concept of depth for shape parameters.

Lastly, the concentration parameter (formally defined, when it exists, as Σ^{-1}), also plays a major role in various domains of statistics, such as graphical models. For example, Gaussian graphical models are used in economics, social sciences or natural sciences to model the relationships between different variables of interest in the form of a graph, for which $(\Sigma^{-1})_{i,j} = 0$ whenever nodes i and j are independent conditionally to all other nodes (see, for example Hojsgaard *et al.* (2012), and references within). Again, no depth function exists for a parameter of that type.

4 Objectives and structure of the thesis

Halfspace depth (and, to a lesser extent, simplicial depth) plays a predominant role in the depth literature due to its geometric nature, able to capture local variations in the distributions. Understanding the geometric properties of a depth function as well as those of its associated depth regions proves crucial in assessing the existence of depth-based medians (that is, a value of the parameter with maximal depth) as well as in designing efficient algorithms to compute them.

The main hurdles in providing meaningful depth notions for dispersion parameters are twofold. First, the set of (scatter, shape or volatility) parameters is of an intricate geometric nature. Those sets can be formally defined as curved Riemannian manifolds. Contrary to the location case, the non-flat structure of these sets makes the study and definition of depths not straightforward. Second, for location, there exists a confounding between the sample space of observations and parameter space (both are $\mathbb{R}^k$). This allows to consider each observation as a potential location parameter and to construct depth notions out of this "dual" nature. This is not so in the dispersion setting, where the parameter sets – formally, $\mathcal{P}_k$, the set of

$k \times k$ positive definite matrices for Σ and Σ^{-1}, and $\mathcal{V}_{S,k}$, the subset of $\mathcal{P}_k$ of matrices satisfying $S(\Sigma) = 1$ for $V_{S,k}$, where S is a scale functional (formally defined later) – are distinct from $\mathbb{R}^k$.

The goal of this manuscript is to develop and study depth notions in, broadly, the dispersion setting, that is, for scatter, shape or concentration matrices. This thesis consists (besides this general Introduction) in two chapters that consider two different approaches. Each chapter has been written in such a way that it can be read independently and provides a self-contained study. In particular, each chapter is based on a paper that has been accepted by or is in revision for an international journal but has been rewritten and rearranged to suit the more detailed exposition of this manuscript.

The **first chapter** builds on the scatter depth concept from Chen *et al.* (2017) and Zhang (2002) and provides generalised depth functions for scatter, shape and concentration parameters. It focuses, in particular, on the properties of the proposed depths and of their depth regions. Of a halfspace nature, the concept requires minimal assumptions and avoids elliptical or absolute continuity constraints. The detailed study requires considering Frobenius and Riemannian topology on $\mathcal{P}_k$ and $\mathcal{V}_{S,k}$. In the spirit of Zuo & Serfling (2000), a paradigmatic approach to scatter depth is also developed and all depth functions are illustrated with simulated and real-data examples.

The **second chapter** introduces a depth function for shape matrices that is of a sign nature. In particular, it involves data points only through their directions from the center of the distribution. The resulting depth median (the deepest matrix) plays the role of estimate of shape and is a depth-based version of the celebrated M-estimator of shape from Tyler (1987). The terminology of Tyler shape depth is therefore used. Properties of the depth are studied. They include invariance, quasi-concavity, and continuity of the depth function, as well as existence and almost sure consistency of a shape median and Fisher consistency in the elliptical setting. Depth-based hypothesis testing for shape parameters is also investigated and the robustness properties of these tests are assessed. Again, practical illustrations are provided throughout.

Finally, a short conclusion closes this thesis, while an Appendix collects all proofs from both chapters.

Chapter I

Halfspace depths for scatter, concentration and shape matrices

Halfspace Depths for Scatter, Concentration and Shape Matrices[1]

Abstract We propose halfspace depth concepts for scatter, concentration and shape matrices. For scatter matrices, our concept is similar to those from Chen *et al.* (2017) and Zhang (2002). Rather than focusing, as in these earlier works, on deepest scatter matrices, we thoroughly investigate the properties of the proposed depth and of the corresponding depth regions. We do so under minimal assumptions and, in particular, we do not restrict to elliptical distributions nor to absolutely continuous distributions. Interestingly, fully understanding scatter halfspace depth requires considering different geometries/topologies on the space of scatter matrices. We also discuss, in the spirit of Zuo & Serfling (2000), the structural properties a scatter depth should satisfy, and investigate whether or not these are met by scatter halfspace depth. Companion concepts of depth for concentration matrices and shape matrices are also proposed and studied. We show the practical relevance of the depth concepts considered in a real-data example from finance.

Keywords: Curved parameter space; Elliptical distributions; Robustness; Scatter matrices; Shape matrices; Statistical Depth.

[1]This chapter is a joint work with Davy Paindaveine. The manuscript has been accepted for publication, although in a different form, in *The Annals of Statistics*.

1 Introduction

Statistical depth measures the centrality of a given location in $\mathbb{R}^k$ with respect to a sample of k-variate observations, or, more generally, with respect to a probability measure P over $\mathbb{R}^k$. The most famous depths include the halfspace depth (Tukey, 1975), the simplicial depth (Liu, 1990), the spatial depth (Vardi & Zhang, 2000) and the projection depth (Zuo, 2003b), see the Introduction to this thesis above for more details. In the last decade, depth has also known much success in functional data analysis, where it measures the centrality of a function with respect to a sample of functional data. Some instances are the band depth (López-Pintado & Romo, 2009), the functional halfspace depth (Claeskens *et al.*, 2014) and the functional spatial depth (Chakraborty & Chaudhuri, 2014). The large variety of available depths made it necessary to introduce an axiomatic approach identifying the most desirable properties of a depth function; see Zuo & Serfling (2000) in the multivariate case and Nieto-Reyes & Battey (2016) in the functional one.

Statistical depth provides a center-outward ordering of the observations that allows to tackle in a robust and nonparametric way a broad range of inference problems; see Liu *et al.* (1999). For most depths, the deepest point is a robust location functional that extends the univariate median to the multivariate or functional setups; see, in particular, Cardot *et al.* (2017) for a recent work on the functional spatial median. Beyond the median, depth plays a key role in the classical problem of defining multivariate quantiles; see, e.g., Hallin *et al.* (2010) or Serfling (2010). In line with this, the collections of locations in $\mathbb{R}^k$ whose depth does not exceed a given level are sometimes called *quantile regions*; see, e.g., He & Einmahl (2017) in a multivariate extreme value theory framework. In the functional case, the quantiles in Chaudhuri (1996) may be seen as those associated with functional spatial depth; see Chakraborty & Chaudhuri (2014). Both in the multivariate and functional cases, supervised classification and outlier detection are standard applications of depth; we refer, e.g., to Cuevas *et al.* (2007), Paindaveine & Van Bever (2015), Dang & Serfling (2010), Hubert *et al.* (2015) and to the references therein.

In Mizera (2002), statistical depth was extended to a virtually arbitrary parametric framework. In a generic parametric model indexed by an ℓ-dimensional parameter ϑ, the resulting *tangent depth* $D_{P_n}(\vartheta_0)$ measures how appropriate a parameter value ϑ_0 is, with respect to the empirical measure P_n of a sample of k-variate observations $X_1, \ldots, X_n$ at hand, as one could alternatively do based on the likelihood $L_{P_n}(\vartheta_0)$. Unlike the MLE of ϑ, the depth-based estimator maximising $D_{P_n}(\vartheta)$ is robust under mild conditions; see Section 4 of Mizera (2002). The construction, that for linear regression provides the Rousseeuw & Hubert (1999) depth, proved useful in various contexts. However, tangent depth requires evaluating the halfspace depth of a given location in $\mathbb{R}^\ell$, hence can only deal with low-dimensional parameters. In particular, tangent depth cannot cope with covariance or scatter matrix parameters ($\ell = k(k+1)/2$), unless k is as small as 2 or 3.

The crucial role played by scatter matrices in multivariate statistics, however, makes it highly desirable to have a satisfactory depth for such parameters, as phrased by Serfling (2004b), that calls for an extension of the Mizera & Müller (2004) location-scale depth concept into a location-scatter one. While computational issues prevent from basing this extension on tangent depth, a more ad hoc approach such as the one proposed in Zhang (2002) is suitable. Recently, another concept of scatter depth, that is very close in spirit to the one from Zhang (2002), was introduced in Chen *et al.* (2017). Both proposals dominate tangent depth in the sense that, for k-variate observations, they rely on projection pursuit in $\mathbb{R}^k$ rather than in $\mathbb{R}^{k(k+1)/2}$, which allowed Chen *et al.* (2017) to consider their depth even in high dimensions, under, e.g., sparsity assumptions. Both works, however, mainly focus on asymptotic, robustness and/or minimax convergence properties of the sample deepest scatter matrix. The properties of these scatter depths thus remain largely unknown, which severely affects the interpretation of the sample concepts.

In the present chapter, we consider a concept of halfspace depth for scatter matrices that is close to the Zhang (2002) and Chen *et al.* (2017) ones. Unlike these previous works, however, we thoroughly study the properties of the scatter depth and of the corresponding depth regions. We do so under minimal assumptions and, in particular, we do not restrict to elliptical distributions nor to absolutely continuous distributions. Interestingly, fully understanding scatter halfspace depth requires considering different geometries/topologies on the space of scatter matrices. Like Donoho & Gasko (1992) and Rousseeuw & Ruts (1999) did for location halfspace depth, we study continuity and quasi-concavity properties of scatter halfspace depth, as well as the boundedness, convexity and compacity properties of the corresponding depth regions. Existence of a deepest halfspace scatter matrix, which is not guaranteed a priori, is also investigated. We further discuss, in the spirit of Zuo & Serfling (2000), the structural properties a scatter depth should satisfy and we investigate whether or not these are met by scatter halfspace depth. Moreover, companion concepts of depth for concentration matrices and shape matrices are proposed and studied. To the best of our knowledge, our results are the first providing structural and topological properties of depth regions outside the classical location framework. Throughout, numerical results illustrate our theoretical findings. Finally, we show the practical relevance of the depth concepts considered in a real-data example from finance.

The outline of the chapter is as follows. In Section 2, we define scatter halfspace depth and investigate its affine invariance and uniform consistency. We also obtain explicit expressions of this depth for two distributions we will use as running examples throughout this chapter. In Section 3, we derive the properties of scatter halfspace depth and scatter halfspace depth regions when considering the Frobenius topology on the space of scatter matrices, whereas we do the same for the geodesic topology in Section 4. In Section 5, we identify the desirable properties a generic scatter depth should satisfy and investigate whether or not these are met by scatter halfspace depth. In Sections 6 and 7, we extend this depth to concentration and shape matrices, respectively. In Section 8, we treat a real-data example from finance. Numerical results are provided throughout the chapter to illustrate the various theorems, while the global Appendix to this thesis collects all proofs from this manuscript.

Before proceeding, we list here, for the sake of convenience, some notation to be used throughout. The collection of $k \times k$ matrices, $k \times k$ invertible matrices, and $k \times k$ symmetric matrices will be denoted as $\mathcal{M}_k$, GL_k, and $\mathcal{S}_k$, respectively (all matrices in this manuscript are real matrices). The identity matrix in $\mathcal{M}_k$ will be denoted as I_k. For any $A \in \mathcal{M}_k$, $\mathrm{diag}(A)$ will stand for the k-vector collecting the diagonal entries of A, whereas, for any k-vector v, $\mathrm{diag}(v)$ will stand for the diagonal matrix such that $\mathrm{diag}(\mathrm{diag}(v)) = v$.

For $p \geq 2$ square matrices $A_1, \ldots, A_p$, $\text{diag}(A_1, \ldots, A_p)$ will stand for the block-diagonal matrix with diagonal blocks $A_1, \ldots, A_p$. Any matrix A in $\mathcal{S}_k$ can be diagonalised into $A = O\, \text{diag}(\lambda_1(A), \ldots, \lambda_k(A))\, O'$, where $\lambda_1(A) \geq \ldots \geq \lambda_k(A)$ are the eigenvalues of A and where the columns of the $k \times k$ orthogonal matrix $O = (v_1(A), \ldots, v_k(A))$ are corresponding unit eigenvectors (as usual, eigenvectors, and possibly eigenvalues, are only partly identified, but this will not play a role in the sequel). The spectral interval of A is $\text{Sp}(A) := [\lambda_k(A), \lambda_1(A)]$. For any mapping $f : \mathbb{R} \to \mathbb{R}$, we let $f(A) = O\, \text{diag}(f(\lambda_1(A)), \ldots, f(\lambda_k(A)))\, O'$. If Σ is a *scatter matrix*, in the sense that Σ belongs to the collection $\mathcal{P}_k$ of symmetric and positive definite $k \times k$ matrices, then this defines $\log(\Sigma)$ and Σ^t for any $t \in \mathbb{R}$. In particular, $\Sigma^{1/2}$ is the unique $A \in \mathcal{P}_k$ such that $\Sigma = AA'$, and $\Sigma^{-1/2}$ is the inverse of this symmetric and positive definite square root. Throughout, T will denote a location functional, that is, a function mapping a probability measure P to a real k-vector T_P. A location functional T is affine-equivariant if $T_{P_{A,b}} = AT_P + b$ for any $A \in GL_k$ and $b \in \mathbb{R}^k$, where the probability measure $P_{A,b}$ is the distribution of $AX + b$ when X has distribution P. A much weaker equivariance concept is centro-equivariance, for which $T_{P_{A,b}} = AT_P + b$ is imposed for $A = -I_k$ and $b = 0$ only. For a probability measure P over $\mathbb{R}^k$ and a location functional T, we will let $\alpha_{P,T} := \min(s_{P,T}, 1 - s_{P,T})$, where $s_{P,T} := \sup_{u \in \mathcal{S}^{k-1}} P[\{x \in \mathbb{R}^k : u'(x - T_P) = 0\}]$ involves the unit sphere $\mathcal{S}^{k-1} := \{x \in \mathbb{R}^k : \|x\|^2 = x'x = 1\}$ of $\mathbb{R}^k$. We will say that P is *smooth at* $\theta(\in \mathbb{R}^k)$ if the P-probability of any hyperplane of $\mathbb{R}^k$ containing θ is zero and that it is *smooth* if it is smooth at any θ. Finally, $\stackrel{\mathcal{D}}{=}$ will denote equality in distribution.

2 Scatter halfspace depth

We start by recalling the classical concept of location halfspace depth. To do so, let P be a probability measure over $\mathbb{R}^k$ and X be a random k-vector with distribution P, which allows us throughout to write $P[X \in B]$ instead of $P[B]$ for any k-Borel set B. The *location halfspace depth* of $\theta(\in \mathbb{R}^k)$ with respect to P is then

$$HD_P^{\text{loc}}(\theta) := \inf_{u \in \mathcal{S}^{k-1}} P[u'(X - \theta) \geq 0].$$

The corresponding *depth regions* $R_P^{\text{loc}}(\alpha) := \{\theta \in \mathbb{R}^k : HD_P^{\text{loc}}(\theta) \geq \alpha\}$ form a nested family of closed convex subsets of $\mathbb{R}^k$. The innermost depth region, namely $M_P^{\text{loc}} := \{\theta \in \mathbb{R}^k : HD_P^{\text{loc}}(\theta) = \max_{\eta \in \mathbb{R}^k} HD_P^{\text{loc}}(\eta)\}$ (the maximum always exists; see, e.g., Proposition 7 in Rousseeuw & Ruts, 1999), is a set-valued location functional. When a unique representative of M_P^{loc} is needed, it is customary to consider the *Tukey median* θ_P of P, that is defined as the barycenter of M_P^{loc}. The Tukey median has maximal depth (which follows from the convexity of M_P^{loc}) and is an affine-equivariant location functional.

In this chapter, for a location functional T, we define the *T-scatter halfspace depth* of $\Sigma(\in \mathcal{P}_k)$ with respect to P as

$$HD_{P,T}^{\text{sc}}(\Sigma) := \inf_{u \in \mathcal{S}^{k-1}} \min\Big(P\big[|u'(X - T_P)| \leq \sqrt{u'\Sigma u}\,\big], \\ P\big[|u'(X - T_P)| \geq \sqrt{u'\Sigma u}\,\big]\Big). \tag{2.1}$$

This extends to a probability measure with arbitrary location the centered matrix depth concept from Chen *et al.* (2017). If P is smooth, then the depth in (2.1) is also equivalent to the (Tukey version of) the dispersion depth introduced in Zhang (2002), but for the fact that the latter, in the spirit of projection depth, involves centering through a univariate location functional (both Zhang (2002) and Chen *et al.* (2017) also propose bypassing centering through a pairwise difference approach that will be discussed in Section 5). While they were not considered in these prior works, it is of interest to introduce the corresponding depth regions

$$R^{\rm sc}_{P,T}(\alpha) := \big\{\Sigma \in \mathcal{P}_k : HD^{\rm sc}_{P,T}(\Sigma) \geq \alpha\big\}, \quad \alpha \geq 0. \tag{2.2}$$

We will refer to $R^{\rm sc}_{P,T}(\alpha)$ as the *order-*α *(*T*-scatter halfspace) depth region of* P. Obviously, one always has $R^{\rm sc}_{P,T}(0) = \mathcal{P}_k$. Clearly, the concepts in (2.1)-(2.2) give practitioners the flexibility to freely choose the location functional T; numerical results below, however, will focus on the depth $HD^{\rm sc}_P(\Sigma)$ and on the depth regions $R^{\rm sc}_P(\alpha)$ based on the Tukey median θ_P, that is the natural location functional whenever halfspace depth objects are considered.

To get a grasp of the scatter depth $HD^{\rm sc}_P(\Sigma)$, it is helpful to start with the univariate case $k = 1$. There, the location halfspace deepest region is the "median interval" $M^{\rm loc}_P = \arg\max_{\theta\in\mathbb{R}} \min(P[X \leq \theta], P[X \geq \theta])$ and the Tukey median θ_P, that is, the midpoint of $M^{\rm loc}_P$, is the usual representative of the univariate median. The scatter halfspace deepest region is then the median interval $M^{\rm sc}_P := \arg\max_{\Sigma\in\mathbb{R}^+_0} \min(P[(X-\theta_P)^2 \leq \Sigma], P[(X-\theta_P)^2 \geq \Sigma])$ of $(X-\theta_P)^2$; call it the *median squared deviation* interval $\mathcal{I}_{\rm MSD}[X]$ (or $\mathcal{I}_{\rm MSD}[P]$) of $X \sim P$. Below, parallel to what is done for the median, $\mathrm{MSD}[X]$ (or $\mathrm{MSD}[P]$) will denote the midpoint of this MSD interval. In particular, if $\mathcal{I}_{\rm MSD}[P]$ is a singleton, then scatter halfspace depth is uniquely maximised at $\Sigma = \mathrm{MSD}[P] = (\mathrm{MAD}[P])^2$, where $\mathrm{MAD}[P]$ denotes the median absolute deviation of P. Obviously, the depth regions $R^{\rm sc}_P(\alpha)$ form a family of nested intervals, $[\Sigma^-_\alpha, \Sigma^+_\alpha]$ say, included in $\mathcal{P}_1 = \mathbb{R}^+_0$. It is easy to check that, if P is symmetric about zero with an invertible cumulative distribution function F and if T is centro-equivariant, then

$$HD^{\rm sc}_P(\Sigma) = HD^{\rm sc}_{P,T}(\Sigma) = 2\min\big(F(\sqrt{\Sigma}) - \tfrac{1}{2}, 1 - F(\sqrt{\Sigma})\big) \quad \text{and} \tag{2.3}$$

$$R^{\rm sc}_P(\alpha) = R^{\rm sc}_{P,T}(\alpha) = \big[(F^{-1}(\tfrac{1}{2}+\tfrac{\alpha}{2}))^2, (F^{-1}(1-\tfrac{\alpha}{2}))^2\big]. \tag{2.4}$$

This is compatible with the fact that the maximal value of $\Sigma \mapsto HD^{\rm sc}_P(\Sigma)$ (that is equal to 1/2) is achieved at $\Sigma = (\mathrm{MAD}[P])^2$ only.

For $k > 1$, elliptical distributions provide an important particular case. We will say that $P = P^X$ is k-variate elliptical with location $\theta(\in \mathbb{R}^k)$ and scatter $\Sigma(\in \mathcal{P}_k)$ if and only if $X \stackrel{\mathcal{D}}{=} \theta + \Sigma^{1/2}Z$, where $Z = (Z_1,\ldots,Z_k)'$ is (i) spherically symmetric about the origin of $\mathbb{R}^k$ (that is, $OZ \stackrel{\mathcal{D}}{=} Z$ for any $k\times k$ orthogonal matrix O) and is (ii) standardised in such a way that $\mathrm{MSD}[Z_1] = 1$ (one then has $T_P = \theta$ for any affine-equivariant location functional T). Denoting by Φ the cumulative distribution function of the standard normal, the k-variate normal distribution with location zero and scatter I_k is then the distribution of $X := W/b$, where $b := \Phi^{-1}(\frac{3}{4})$ and W is a standard normal random k-vector. In this Gaussian case, we obtain

$$\begin{aligned} HD^{\rm sc}_{P,T}(\Sigma) &= \inf_{u\in\mathcal{S}^{k-1}} \min\big(P\big[|u'X| \leq \sqrt{u'\Sigma u}\,\big], P\big[|u'X| \geq \sqrt{u'\Sigma u}\,\big]\big) \\ &= 2\min\Big(\Phi\big(b\lambda_k^{1/2}(\Sigma)\big) - \tfrac{1}{2}, 1 - \Phi\big(b\lambda_1^{1/2}(\Sigma)\big)\Big). \end{aligned} \tag{2.5}$$

One can check directly that $HD^{\rm sc}_{P,T}(\Sigma) \leq HD^{\rm sc}_{P,T}(I_k) = 1/2$, with equality if and only if Σ coincides with the "true" scatter matrix I_k (we refer to Theorem 5.1 for a more general result). Also, Σ belongs to the depth region $R^{\rm sc}_{P,T}(\alpha)$ if and only if $\mathrm{Sp}(\Sigma) \subset [(\frac{1}{b}\Phi^{-1}(\frac{1}{2}+\frac{\alpha}{2}))^2, (\frac{1}{b}\Phi^{-1}(1-\frac{\alpha}{2}))^2]$.

Provided that the location functional used is affine-equivariant, extension to an arbitrary multinormal is based on the following affine-invariance result, which ensures in particular that scatter halfspace depth will not be affected by possible changes in the marginal measurement units (a similar result is stated in Zhang (2002) for the dispersion depth concept considered there).

Theorem 2.1. Let T be an affine-equivariant location functional. Then, (i) scatter halfspace depth is affine-invariant in the sense that, for any probability measure P over $\mathbb{R}^k$, $\Sigma \in \mathcal{P}_k$, $A \in GL_k$ and $b \in \mathbb{R}^k$, we have $HD^{\rm sc}_{P_{A,b},T}(A\Sigma A') = HD^{\rm sc}_{P,T}(\Sigma)$, where $P_{A,b}$ is as defined on page 16. Consequently, (ii) the regions $R^{\rm sc}_{P,T}(\alpha)$ are affine-equivariant, in the sense that, for any probability measure P over $\mathbb{R}^k$, $\alpha \geq 0$, $A \in GL_k$ and $b \in \mathbb{R}^k$, we have $R^{\rm sc}_{P_{A,b},T}(\alpha) = AR^{\rm sc}_{P,T}(\alpha)A'$.

This result readily entails that if P is the k-variate normal with location θ_0 and scatter Σ_0, then, provided that T is affine-equivariant,

$$HD^{\rm sc}_{P,T}(\Sigma) = 2\min\Big(\Phi\big(b\lambda_k^{1/2}(\Sigma_0^{-1}\Sigma)\big) - \tfrac{1}{2}, 1 - \Phi\big(b\lambda_1^{1/2}(\Sigma_0^{-1}\Sigma)\big)\Big) \tag{2.6}$$

and $R^{\rm sc}_{P,T}(\alpha)$ is the collection of scatter matrices Σ for which the spectral interval satisfies $\mathrm{Sp}(\Sigma_0^{-1}\Sigma) \subset [(\frac{1}{b}\Phi^{-1}(\frac{1}{2}+\frac{\alpha}{2}))^2, (\frac{1}{b}\Phi^{-1}(1-\frac{\alpha}{2}))^2]$. For a non-Gaussian elliptical probability measure P with location θ_0 and scatter Σ_0, it is easy to show that $HD^{\rm sc}_{P,T}(\Sigma)$ will still depend on Σ only through $\lambda_1(\Sigma_0^{-1}\Sigma)$ and $\lambda_k(\Sigma_0^{-1}\Sigma)$.

As already mentioned, we also intend to consider non-elliptical probability measures. A running non-elliptical example will be the one for which P is the distribution of a random vector $X = (X_1, \ldots, X_k)'$ with independent Cauchy marginals. If T is centro-equivariant, then

$$HD^{\rm sc}_{P,T}(\Sigma) = 2\min\Big(\Psi\big(1/\max_s \sqrt{s'\Sigma^{-1}s}\,\big) - \tfrac{1}{2}, 1 - \Psi\big(\sqrt{\max(\mathrm{diag}(\Sigma))}\big)\Big), \tag{2.7}$$

where Ψ is the Cauchy cumulative distribution function and where the maximum in s is over all sign vectors $s = (s_1, \ldots, s_k) \in \{-1, 1\}^k$; see Lemma A.1 in the Appendix for a proof. For $k = 1$, this simplifies to $HD^{\rm sc}_{P,T}(\Sigma) = 2\min\big(\Psi\big(\sqrt{\Sigma}\,\big) - \frac{1}{2}, 1 - \Psi\big(\sqrt{\Sigma}\big)\big)$, which agrees with (2.3). For $k = 2$, we obtain

$$HD^{\rm sc}_{P,T}(\Sigma) = 2\min\Big(\Psi\big(\sqrt{\det(\Sigma)/s_\Sigma}\,\big) - \tfrac{1}{2}, 1 - \Psi\big(\sqrt{\max(\Sigma_{11}, \Sigma_{22})}\big)\Big),$$

where we let $s_\Sigma := \Sigma_{11} + \Sigma_{22} + 2|\Sigma_{12}|$. For a general k, a scatter matrix Σ belongs to $R^{\rm sc}_{P,T}(\alpha)$

if and only if $1/(s'\Sigma^{-1}s) \geq (\Psi^{-1}(\frac{1}{2}+\frac{\alpha}{2}))^2$ for all $s \in \{-1,1\}^k$ and $\Sigma_{\ell\ell} \leq (\Psi^{-1}(1-\frac{\alpha}{2}))^2$ for all $\ell = 1,\ldots,k$. The problem of identifying the scatter matrix achieving maximal depth, if any (existence is not guaranteed), will be considered in Section 4. Figure 1 plots scatter halfspace depth regions in the Gaussian and independent Cauchy cases above. Examples involving distributions that are not absolutely continuous with respect to the Lebesgue measure will be considered in the next sections.

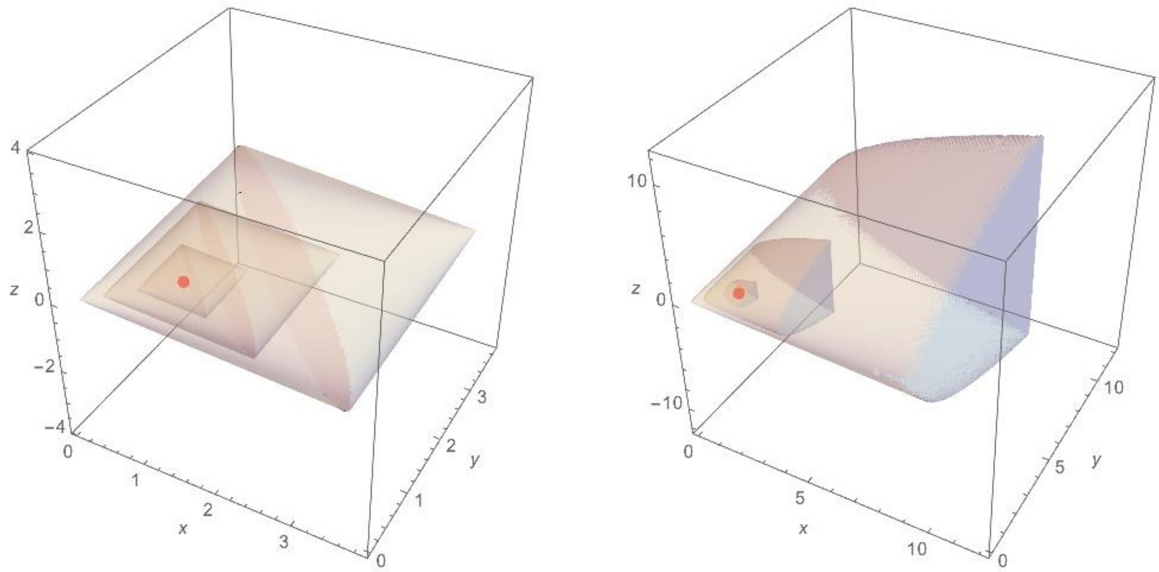

Figure 1: *Level sets of order $\alpha = .2, .3$ and $.4$, for any centro-symmetric T, of $(x,y,z) \mapsto HD^{\rm sc}_{P,T}(\Sigma_{x,y,z})$, where $HD^{\rm sc}_{P,T}(\Sigma_{x,y,z})$ is the T-scatter halfspace depth of $\Sigma_{x,y,z} = \left(\begin{smallmatrix} x & z \\ z & y \end{smallmatrix}\right)$ with respect to two probability measures P, namely the bivariate multinormal distribution with location zero and scatter I_2 (left) and the bivariate distribution with independent Cauchy marginals (right). The red points are those associated with I_2 (left) and $\sqrt{2}I_2$ (right), which are the corresponding deepest scatter matrices (see Sections 4 and 5).*

The expressions for $HD^{\rm sc}_{P,T}(\Sigma)$ obtained in (2.6)-(2.7) above are validated through a Monte Carlo exercise below. Such a numerical validation is justified by the following uniform consistency result; see (6.2) and (6.6) in Donoho & Gasko (1992) for the corresponding location halfspace depth result, and Proposition 2.2(ii) in Zhang (2002) for the dispersion depth concept considered there.

Theorem 2.2. Let P be a smooth probability measure over $\mathbb{R}^k$ and T be a location functional. Let P_n denote the empirical probability measure associated with a random sample of size n from P and assume that $T_{P_n} \to T_P$ almost surely as $n \to \infty$. Then $\sup_{\Sigma \in \mathcal{P}_k} |HD^{\rm sc}_{P_n,T}(\Sigma) - HD^{\rm sc}_{P,T}(\Sigma)| \to 0$ almost surely as $n \to \infty$.

This result applies in particular to the scatter halfspace depth $HD^{\rm sc}_P(\Sigma)$, as the Tukey median is strongly consistent without any assumption on P (for completeness, we show this in the appendix, see Lemma A.5). Of course, the uniform consistency result in Theorem 2.2 holds in particular if P is absolutely continuous with respect to the Lebesgue measure. Inspection of the proof of Theorem 2.2 reveals that the smoothness assumption is only needed to control the estimation of T_P, hence is superfluous when a constant location functional is used. This is relevant when the location is fixed, as in Chen *et al.* (2017).

The Monte Carlo exercise we use to validate these expressions is the following. For each possible combination of $n \in \{100, 500, 2000\}$ and $k \in \{2, 3, 4\}$, we generated $M = 1000$ independent random samples of size n (i) from the k-variate normal distribution with location zero and scatter I_k and (ii) from the k-variate distribution with independent Cauchy marginals. Letting $\Lambda_k^A := \mathrm{diag}(8, I_{k-1})$, $\Lambda_k^B := I_k$, $\Lambda_k^C := \mathrm{diag}(\frac{1}{8}, I_{k-1})$, and

$$O_k := \mathrm{diag}\left(\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}, I_{k-2}\right),$$

we evaluated, in each sample, the depths $HD^{\mathrm{sc}}_{P_n}(\Sigma_\ell)$ of $\Sigma_\ell = O_k \Lambda_k^\ell O_k'$, $\ell = A, B, C$, where P_n denotes the empirical probability measure associated with the k-variate sample of size n at hand (each evaluation of $HD^{\mathrm{sc}}_{P_n}(\cdot)$ is done by approximating the infimum in $u \in \mathcal{S}^{k-1}$ by a minimum over $N = 10000$ directions randomly sampled from the uniform distribution over $\mathcal{S}^{k-1}$). For each n, k, and each underlying distribution (multinormal or independent Cauchy), Figure 2 reports boxplots of the corresponding M values of $HD^{\mathrm{sc}}_{P_n}(\Sigma_\ell)$, $\ell = A, B, C$. Clearly, the results support the theoretical depth expressions obtained in (2.6)-(2.7), as well as the consistency result in Theorem 2.2 (the bias for $HD^{\mathrm{sc}}_{P_n}(\Sigma_B)$ in the Gaussian case is explained by the fact that, as we have seen above, Σ_B maximises $HD^{\mathrm{sc}}_{P}(\Sigma)$, with a maximal depth value equal to 1/2).

3 Frobenius topology

Our investigation of the further structural properties of the scatter halfspace depth $HD^{\mathrm{sc}}_{P,T}(\Sigma)$ and of the corresponding depth regions $R^{\mathrm{sc}}_{P,T}(\alpha)$ depends on the topology that is considered on $\mathcal{P}_k$. In this section, we focus on the topology induced by the *Frobenius metric space* $(\mathcal{P}_k, d_F)$, where $d_F(\Sigma_a, \Sigma_b) = \|\Sigma_b - \Sigma_a\|_F$ is the distance on $\mathcal{P}_k$ that is inherited from the Frobenius norm $\|A\|_F = \sqrt{\mathrm{tr}[AA']}$ on $\mathcal{M}_k$. The resulting *Frobenius topology* (or simply *F-topology*), generated by the F-balls $B_F(\Sigma_0, r) := \{\Sigma \in \mathcal{P}_k : d_F(\Sigma, \Sigma_0) < r\}$ with center Σ_0 and radius r, gives a precise meaning to what we call below *F-continuous* functions on $\mathcal{P}_k$, *F-open/F-closed* subsets of $\mathcal{P}_k$, etc. We then have the following result.

Theorem 3.1. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, (i) $\Sigma \mapsto HD^{\mathrm{sc}}_{P,T}(\Sigma)$ is upper F-semicontinuous on $\mathcal{P}_k$, so that (ii) the depth region $R^{\mathrm{sc}}_{P,T}(\alpha)$ is F-closed for any $\alpha \geq 0$. (iii) If P is smooth at T_P, then $\Sigma \mapsto HD^{\mathrm{sc}}_{P,T}(\Sigma)$ is F-continuous on $\mathcal{P}_k$.

For location halfspace depth, the result corresponding to Theorem 3.1 above was derived in Lemma 6.1 of Donoho & Gasko (1992), where the metric on $\mathbb{R}^k$ is the Euclidean one. The similarity between the location and scatter halfspace depths also extends to the boundedness of depth regions, in the sense that, like location halfspace depth (Proposition 5 in Rousseeuw & Ruts, 1999), the order-α scatter halfspace depth region is bounded if and only if $\alpha > 0$.

Theorem 3.2. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional.

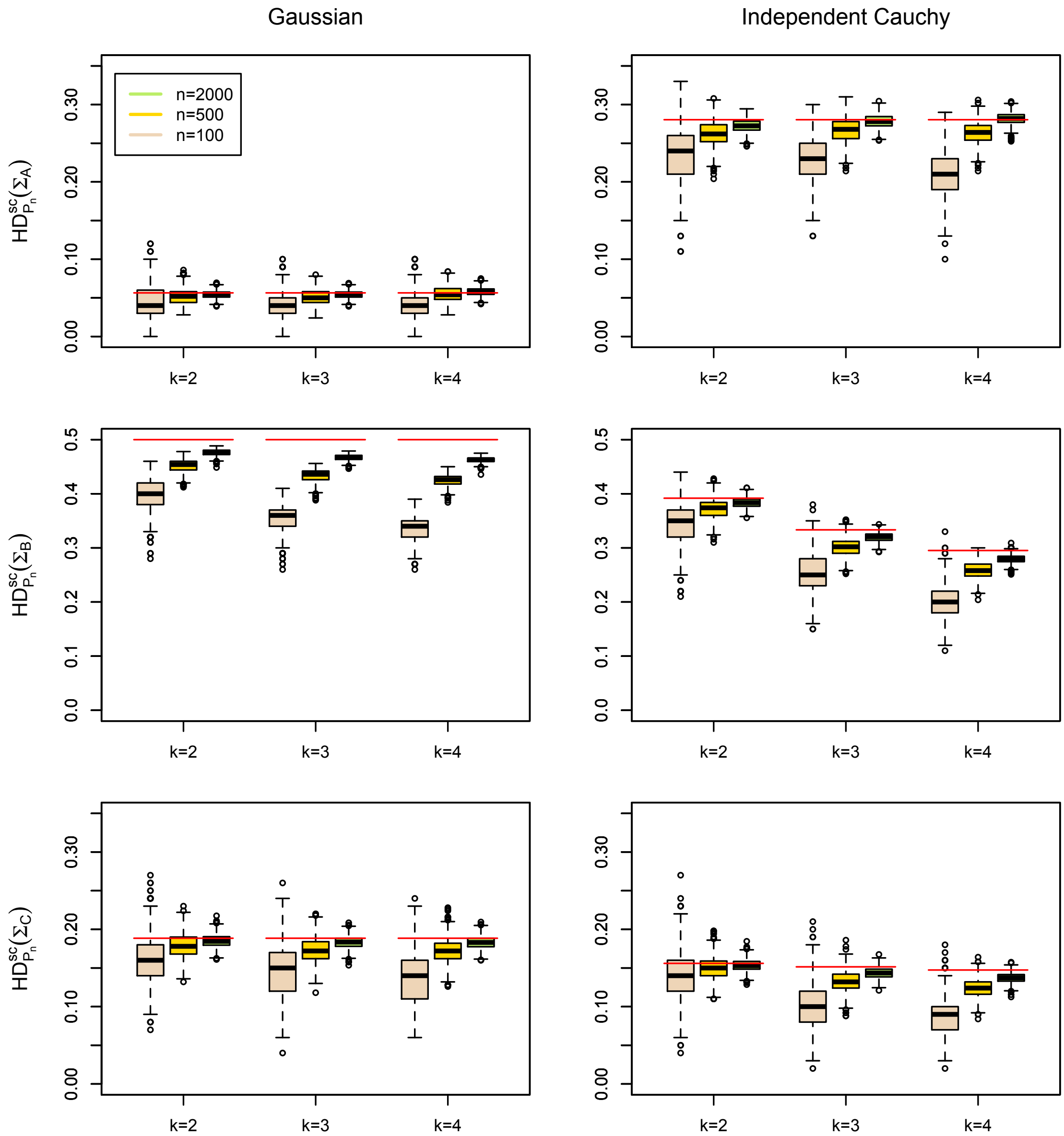

Figure 2: *Boxplots, for various values of n and k, of $HD^{\rm sc}_{P_n}(\Sigma_A)$ (top row), $HD^{\rm sc}_{P_n}(\Sigma_B)$ (middle row) and $HD^{\rm sc}_{P_n}(\Sigma_C)$ (bottom row) based on $M = 1000$ independent random samples of size n from the k-variate multinormal distribution with location zero and scatter I_k (left) or from the k-variate distribution with independent Cauchy marginals (right); we refer to Section 2 for the expressions of Σ_A, Σ_B and Σ_C.*

Then, for any $\alpha > 0$, $R^{\rm sc}_{P,T}(\alpha)$ is F-bounded (that is, it is included, for some $r > 0$, in the F-ball $B_F(I_k, r)$).

This shows that, for any probability measure P, $HD^{\rm sc}_{P,T}(\Sigma)$ goes to zero as $\|\Sigma\|_F \to \infty$. Since $\|\Sigma\|_F \geq \lambda_1(\Sigma)$, this means that explosion of Σ (that is, $\lambda_1(\Sigma) \to \infty$) leads to arbitrarily small depth, which is confirmed in the multinormal case in (2.5). In this Gaussian case, however, implosion of Σ (that is, $\lambda_k(\Sigma) \to 0$) also provides arbitrarily small depth, but this is not captured by the general result in Theorem 3.2 (similar comments can be given for the independent Cauchy example in (2.7)). Irrespective of the topology adopted (so that the F-topology is not to be blamed for this behavior), it is actually possible to have implosion without depth going to zero. We show this by considering the following example. Let $P = (1-s)P_1 + sP_2$, where $s \in (\frac{1}{2}, 1)$, P_1 is the bivariate standard normal and P_2 is the distribution of $\binom{0}{Z}$, where Z is univariate standard normal. Then, it can be showed that, for $\Sigma_n := \left(\begin{smallmatrix} 1/n & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and any centro-equivariant T, we have $HD^{\rm sc}_{P,T}(\Sigma_n) \to 1 - s > 0$ as $n \to \infty$.

In the metric space $(\mathcal{P}_k, d_F)$, any bounded set is also *totally bounded*, that is, can be covered, for any $\varepsilon > 0$, by finitely many balls of the form $B_F(\Sigma, \varepsilon)$. Theorems 3.1-3.2 thus show that, for any $\alpha > 0$, $R^{\rm sc}_{P,T}(\alpha)$ is both F-closed and totally F-bounded. However, since $(\mathcal{P}_k, d_F)$ is not complete, there is no guarantee that these regions are F-compact. Actually, these regions may fail to be F-compact, as we show through the example from the previous paragraph. For any $\alpha \in (0, 1-s)$, the scatter matrix Σ_n belongs to $R^{\rm sc}_{P,T}(\alpha)$ for n large enough. However, the sequence (Σ_n) F-converges to $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, that does not belong to $R^{\rm sc}_{P,T}(\alpha)$ (since it does not even belong to $\mathcal{P}_2$). Since this will also hold for any subsequence of (Σ_n), we conclude that, for $\alpha \in (0, 1-s)$, $R^{\rm sc}_{P,T}(\alpha)$ is not F-compact in this example. This provides a first discrepancy between location and scatter halfspace depths, since location halfspace depth regions associated with a positive order α are always compact.

The lack of compacity of scatter halfspace depth regions may allow for probability measures for which no halfspace deepest scatter exists. This is actually the case in the bivariate mixture example above. There, letting $e_1 = (1, 0)'$ and assuming again that T is centro-equivariant, any $\Sigma \in \mathcal{P}_2$ indeed satisfies $HD^{\rm sc}_{P,T}(\Sigma) \leq P[|e_1'X| \geq \sqrt{e_1'\Sigma e_1}] = P[|X_1| \geq \sqrt{\Sigma_{11}}] = (1 - s)P[|Z| \geq \sqrt{\Sigma_{11}}] < 1 - s = \sup_{\Sigma \in \mathcal{P}_2} HD^{\rm sc}_{P,T}(\Sigma)$, where the last equality follows from the fact that we identified a sequence (Σ_n) such that $HD^{\rm sc}_{P,T}(\Sigma_n) \to 1 - s$. This is again in sharp contrast with the location case, for which a halfspace deepest location always exists; see, e.g., Propositions 5 and 7 in Rousseeuw & Ruts (1999). Identifying sufficient conditions under which a halfspace deepest scatter exists requires considering another topology, namely the *geodesic topology* considered in Section 4 below.

The next result states that scatter halfspace depth is a quasi-concave function, which ensures convexity of the corresponding depth regions; we refer to Proposition 1 (and to its corollary) in Rousseeuw & Ruts (1999) for the corresponding results on location halfspace depth.

Theorem 3.3. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, (i) $\Sigma \mapsto HD^{\rm sc}_{P,T}(\Sigma)$ is quasi-concave, in the sense that, for any $\Sigma_a, \Sigma_b \in \mathcal{P}_k$ and $t \in [0, 1]$, $HD^{\rm sc}_{P,T}(\Sigma_t) \geq \min(HD^{\rm sc}_{P,T}(\Sigma_a), HD^{\rm sc}_{P,T}(\Sigma_b))$, where we let $\Sigma_t := (1-t)\Sigma_a + t\Sigma_b$; (ii) for any $\alpha \geq 0$, $R^{\rm sc}_{P,T}(\alpha)$ is convex.

Strictly speaking, Theorem 3.3 is not directly related to the F-topology considered on $\mathcal{P}_k$. Yet we state the result in this section due to the link between the linear paths $t \mapsto \Sigma_t = (1-t)\Sigma_a + t\Sigma_b$ it involves and the "flat" nature of the F-topology (this link will become clearer below when we will compare with what occurs for the geodesic topology).

Figure 3 plots, for $k = 2, 3$, the graphs of $t \mapsto HD^{\rm sc}_P(\Sigma_t)$ for $\Sigma_t := (1-t)\Sigma_A + t\Sigma_C$, where Σ_A and Σ_C are the scatter matrices considered in the numerical exercise performed in Section 2 and where P is either the k-variate normal distribution with location zero and scatter I_k or the k-variate distribution with independent Cauchy marginals. The same figure also provides the corresponding sample plots, based on a single random sample of size $n = 50$ drawn from each of these two distributions. All plots are compatible with the quasi-concavity result in Theorem 3.3. Figure 3 also illustrates the continuity of $t \mapsto HD^{\rm sc}_{P,T}(\Sigma_t)$ for smooth probability measures P (Theorem 3.1) and shows that continuity may fail to hold in the sample case. Further illustration of Theorem 3.3 will be provided in Figure 4.

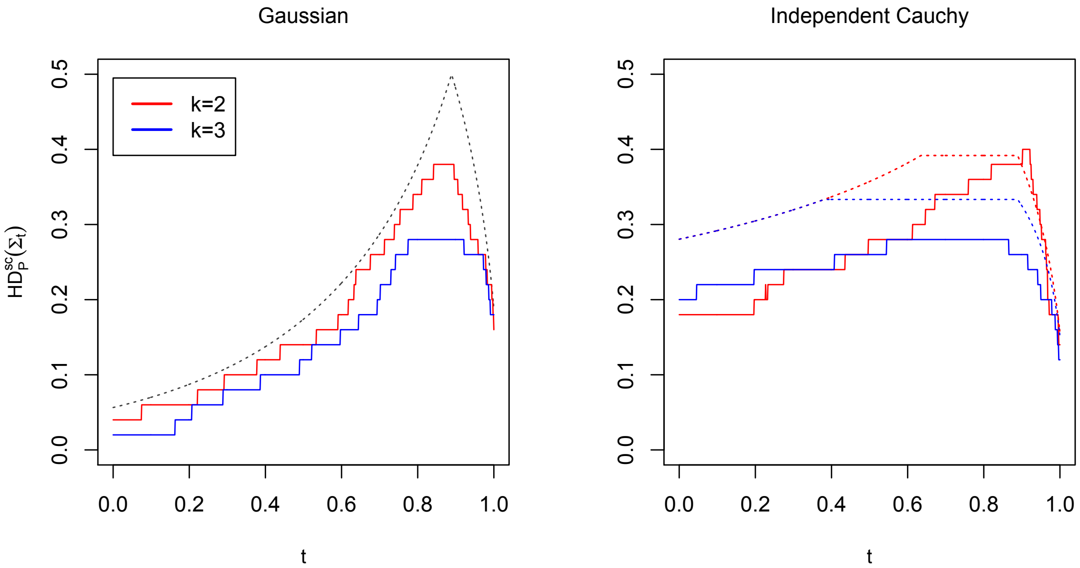

Figure 3: *The dotted curves are the graphs of $t \mapsto HD^{\rm sc}_P(\Sigma_t)$, where $\Sigma_t = (1-t)\Sigma_A + t\Sigma_C$ involves the scatter matrices Σ_A and Σ_C considered in Figure 2, for the k-variate normal distribution with location zero and scatter I_2 (left) and for the k-variate distribution with independent Cauchy marginals (right). The solid curves are associated with a single random sample of size $n = 50$ from the corresponding distributions. Red and blue correspond to the bivariate and trivariate cases, respectively; in the population Gaussian case, the graph of $t \mapsto HD^{\rm sc}_P(\Sigma_t)$ is the same for $k = 2$ and $k = 3$, hence is plotted in black.*

4 Geodesic topology

Equipped with the inner product $< A, B >= {\rm tr}[A'B]$, $\mathcal{M}_k$ is a Hilbert space. The resulting norm and distance are the Frobenius ones considered in the previous section. As an open set in $\mathcal{S}_k$, the parameter space $\mathcal{P}_k$ of interest is a differentiable manifold of dimension $k(k+1)/2$. The corresponding tangent space at Σ, which is isomorphic (via translation) to $\mathcal{S}_k$, can be equipped with the inner product $< A, B > \, = {\rm tr}[\Sigma^{-1}A\Sigma^{-1}B]$. This leads to considering $\mathcal{P}_k$ as a Riemannian manifold, with the metric at Σ given by the differential $ds = \|\Sigma^{-1/2} d\Sigma\, \Sigma^{-1/2}\|_F$; see, e.g., Bhatia (2007). The length of a path $\gamma : [0, 1] \to \mathcal{P}_k$ is then given by

$$L(\gamma) = \int_0^1 \left\| \gamma^{-1/2}(t) \frac{d\gamma(t)}{dt} \gamma^{-1/2}(t) \right\|_F dt.$$

The resulting geodesic distance between $\Sigma_a, \Sigma_b \in \mathcal{P}_k$ is defined as

$$d_g(\Sigma_a, \Sigma_b) := \inf \left\{ L(\gamma) : \gamma \in \mathcal{G}(\Sigma_a, \Sigma_b) \right\} = \| \log(\Sigma_a^{-1/2} \Sigma_b \Sigma_a^{-1/2}) \|_F, \qquad (4.1)$$

where $\mathcal{G}(\Sigma_a, \Sigma_b)$ denotes the collection of paths γ from $\gamma(0) = \Sigma_a$ to $\gamma(1) = \Sigma_b$ (the second equality in (4.1) is Theorem 6.1.6 in Bhatia, 2007). It directly follows from the definition of $d_g(\Sigma_a, \Sigma_b)$ that the geodesic distance satisfies the triangle inequality. Theorem 6.1.6 in Bhatia (2007) also states that all paths γ achieving the infimum in (4.1) provide the same geodesic $\{\gamma(t) : t \in [0,1]\}$ joining Σ_a and Σ_b, and that this geodesic can be parametrised as

$$\gamma(t) = \tilde{\Sigma}_t := \Sigma_a^{1/2}\big(\Sigma_a^{-1/2}\Sigma_b\Sigma_a^{-1/2}\big)^t\Sigma_a^{1/2}, \qquad t \in [0,1]. \tag{4.2}$$

By using the explicit formula in (4.1), it is easy to check that this particular parametrisation of this unique geodesic is natural in the sense that $d_g(\Sigma_a, \tilde{\Sigma}_t) = t d_g(\Sigma_a, \Sigma_b)$ for any $t \in [0,1]$.

Below, we consider the natural topology associated with the metric space $(\mathcal{P}_k, d_g)$, that is, the topology whose open sets are generated by geodesic balls of the form $B_g(\Sigma_0, r) := \{\Sigma \in \mathcal{P}_k : d_g(\Sigma, \Sigma_0) < r\}$. This topology — call it the *geodesic topology*, or simply *g-topology* — defines subsets of $\mathcal{P}_k$ that are g-open, g-closed, g-compact, and functions that are g-semicontinuous, g-continuous, etc. We will say that a subset R of $\mathcal{P}_k$ is g-bounded if and only if $R \subset B_g(I_k, r)$ for some $r > 0$ (we can safely restrict to balls centered at I_k since the triangle inequality guarantees that R is included in a finite-radius g-ball centered at I_k if and only if it is included in a finite-radius g-ball centered at an arbitrary $\Sigma_0 \in \mathcal{P}_k$). A g-bounded subset of $\mathcal{P}_k$ is also totally g-bounded, still in the sense that, for any $\varepsilon > 0$, it can be covered by finitely many balls of the form $B_g(\Sigma, \varepsilon)$; for completeness, we prove this in Lemma A.6 in the Appendix. Since $(\mathcal{P}_k, d_g)$ is complete (see, e.g., Proposition 10 in Bhatia & Holbrook, 2006), a g-bounded and g-closed subset of $\mathcal{P}_k$ is then g-compact.

We omit the proof of the next result as it follows along the exact same lines as the proof of Theorem 3.1, once it is seen that a sequence (Σ_n) converging to Σ_0 in $(\mathcal{P}_k, d_g)$ also converges to Σ_0 in $(\mathcal{P}_k, d_F)$.

Theorem 4.1. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, (i) $\Sigma \mapsto HD^{\rm sc}_{P,T}(\Sigma)$ is upper g-semicontinuous on $\mathcal{P}_k$, so that (ii) the depth region $R^{\rm sc}_{P,T}(\alpha)$ is g-closed for any $\alpha \geq 0$. (iii) If P is smooth at T_P, then $\Sigma \mapsto HD^{\rm sc}_{P,T}(\Sigma)$ is g-continuous on $\mathcal{P}_k$.

The following result uses the notation $s_{P,T} := \sup_{u \in \mathcal{S}^{k-1}} P[u'(X - T_P) = 0]$ and $\alpha_{P,T} := \min(s_{P,T}, 1 - s_{P,T})$ defined in the introduction.

Theorem 4.2. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, for any $\alpha > \alpha_{P,T}$, $R^{\rm sc}_{P,T}(\alpha)$ is g-bounded, hence g-compact (if $s_{P,T} \geq 1/2$, then this result is trivial in the sense that $R^{\rm sc}_{P,T}(\alpha)$ is empty for any $\alpha > \alpha_{P,T}$). In particular, if P is smooth at T_P, then $R^{\rm sc}_{P,T}(\alpha)$ is g-compact for any $\alpha > 0$.

This result complements Theorem 3.2 by showing that implosion always leads to a depth that is smaller than or equal to $\alpha_{P,T}$. In particular, in the multinormal and independent Cauchy examples in Section 2, this shows that both explosion and implosion lead to arbitrarily small depth, whereas Theorem 3.2 was predicting this collapsing for explosion only. Therefore, while the behavior of $HD^{\rm sc}_{P,T}(\Sigma)$ under implosion/explosion of Σ is independent of the topology adopted, the use of the g-topology provides a better understanding of this behavior than the F-topology.

It is not possible to improve the result in Theorem 4.2, in the sense that $R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$ may fail to be g-bounded. For instance, consider the probability measure P over $\mathbb{R}^2$ putting probability mass $1/6$ on each of the six points $(0, \pm 1/2)$ and $(\pm 2, \pm 2)$, and let T be a centro-equivariant location functional. Clearly, $\alpha_{P,T} = s_{P,T} = 1/3$. Now, letting $\Sigma_n := \left(\begin{smallmatrix} 1/n & 0 \\ 0 & 1 \end{smallmatrix}\right)$, we have $P[|u'\Sigma_n^{-1/2}X| \leq 1] \geq 1/3$ and $P[|u'\Sigma_n^{-1/2}X| \geq 1] \geq 1/3$ for any $u \in \mathcal{S}^1$ (here, X is a random vector with distribution P), which entails that

$$HD^{\mathrm{sc}}_{P,T}(\Sigma_n) = \inf_{u \in \mathcal{S}^1} \min\left(P\left[|u'X| \leq \sqrt{u'\Sigma_n u}\,\right], P\left[|u'X| \geq \sqrt{u'\Sigma_n u}\,\right]\right)$$

$$= \inf_{u \in \mathcal{S}^1} \min\left(P[|u'\Sigma_n^{-1/2}X| \leq 1], P[|u'\Sigma_n^{-1/2}X| \leq 1]\right) \geq \frac{1}{3} = \alpha_{P,T},$$

so that $\Sigma_n \in R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$ for any n. Since $d_g(\Sigma_n, I_2) \to \infty$, $R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$ is indeed g-unbounded.

An important benefit of working with the g-topology is that, unlike the F-topology, it allows to show that, under mild assumptions, a halfspace deepest scatter does exist. More precisely, we have the following result.

Theorem 4.3. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Assume that $R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$ is non-empty. Then, $\alpha_{*P,T} := \sup_{\Sigma \in \mathcal{P}_k} HD^{\mathrm{sc}}_{P,T}(\Sigma) = HD^{\mathrm{sc}}_{P,T}(\Sigma_*)$ for some $\Sigma_* \in \mathcal{P}_k$.

In particular, this result shows that for any probability measure P that is smooth at T_P, there exists a halfspace deepest scatter Σ_*. For the k-variate multinormal distribution with location zero and scatter I_k (and any centro-equivariant T), we already stated in Section 2 that $\Sigma \mapsto HD^{\mathrm{sc}}_{P,T}(\Sigma)$ is uniquely maximised at $\Sigma_* = I_k$, with a corresponding maximal depth equal to $1/2$. The next result identifies the halfspace deepest scatter (and the corresponding maximal depth) in the independent Cauchy case.

Theorem 4.4. Let P be the k-variate probability measure with independent Cauchy marginals and let T be a centro-equivariant location functional. Then, $\Sigma \mapsto HD^{\mathrm{sc}}_{P,T}(\Sigma)$ is uniquely maximised at $\Sigma_* = \sqrt{k} I_k$, and the corresponding maximal depth is $HD^{\mathrm{sc}}_{P,T}(\Sigma_*) = \frac{2}{\pi} \arctan\left(k^{-1/4}\right)$.

For $k = 1$, the Cauchy distribution in this result is symmetric (hence, elliptical) about zero, which is compatible with the maximal depth being equal to $1/2$ there (Theorem 5.1 below shows that the maximal depth for absolutely continuous elliptical distributions is always equal to $1/2$). For larger values of k, however, this provides an example where the maximal depth is strictly smaller than $1/2$. Interestingly, this maximal depth goes (monotonically) to zero as $k \to \infty$. Note that, for the same distribution, location halfspace depth has, irrespective of k, maximal value $1/2$ (this follows, e.g., from Lemma 1 and Theorem 1 in Rousseeuw & Struyf, 2004).

In general, the halfspace deepest scatter Σ_* is not unique. This is typically the case for empirical probability measures P_n (note that the existence of a halfspace deepest scatter in the empirical case readily follows from the fact that $HD^{\mathrm{sc}}_{P_n,T}(\Sigma)$ takes its values in $\{\ell/n : \ell = 0, 1, \ldots, n\}$). For several purposes, it is needed to identify a unique representative of the halfspace deepest scatters, that would play a similar role for scatter as the one played by the Tukey median for location. To this end, one may consider here a *center of mass*, that is, a scatter matrix of the form

$$\Sigma_{P,T} := \arg\min_{\Sigma\in\mathcal{P}_k} \int_{R^{\rm sc}_{P,T}(\alpha_{*P,T})} d_g^2(m,\Sigma)\,dm, \tag{4.3}$$

where dm is a mass distribution on $R^{\rm sc}_{P,T}(\alpha_{*P,T})$ with total mass one (the natural choice being the uniform over $R^{\rm sc}_{P,T}(\alpha_{*P,T})$). This is a suitable solution if $R^{\rm sc}_{P,T}(\alpha_{*P,T})$ is g-bounded (hence, g-compact), since Cartan (1929) showed that, in a simply connected manifold with non-positive curvature (as $\mathcal{P}_k$), every compact set has a unique center of mass; see also Proposition 60 in Berger (2003). Convexity of $R^{\rm sc}_{P,T}(\alpha_{*P,T})$ then ensures that $\Sigma_{P,T}$ has maximal depth. Like for location, this choice of $\Sigma_{P,T}$ as a representative of the deepest scatters guarantees affine equivariance (in the sense that $\Sigma_{P_{A,b},T} = A\Sigma_{P,T}A'$ for any $A \in GL_k$ and any $b \in \mathbb{R}^k$), provided that T itself is affine-equivariant. An alternative approach is to consider the scatter matrix $\Sigma_{P,T}$ whose vectorised form $\operatorname{vec}\Sigma_{P,T}$ is the barycenter of $\operatorname{vec} R^{\rm sc}_{P,T}(\alpha_{*P,T})$. While this is a more practical solution for scatter matrices, the non-flat nature of some of the parameter spaces in Section 7 will require the more involved, manifold-type, approach in (4.3).

As a final comment related to Theorem 4.3, note that if $R^{\rm sc}_{P,T}(\alpha_{P,T})$ is empty, then it may actually be so that no halfspace deepest scatter does exist. An example is provided by the bivariate mixture distribution P in Section 3. There, we saw that, for any centro-equivariant T, no halfspace deepest scatter does exist, which is compatible with the fact that, for any Σ, $HD^{\rm sc}_{P,T}(\Sigma) < 1 - s = \alpha_{P,T}$, so that $R^{\rm sc}_{P,T}(\alpha_{P,T})$ is empty.

5 An axiomatic approach for scatter depth

Building on the properties derived in Liu (1990) for simplicial depth, Zuo & Serfling (2000) introduced an axiomatic approach suggesting that a generic location depth $D^{\rm loc}_P(\,\cdot\,) : \mathbb{R}^k \to [0,1]$ should satisfy the following properties: (P1) affine invariance, (P2) maximality at the symmetry center (if any), (P3) monotonicity relative to any deepest point, and (P4) vanishing at infinity. Without entering into details, these properties are to be understood as follows: (P1) means that $D^{\rm loc}_{P_{A,b}}(A\theta + b) = D^{\rm loc}_P(\theta)$ for any $A \in GL_k$ and $b \in \mathbb{R}^k$, where $P_{A,b}$ is as defined on page 16; (P2) states that if P is symmetric (in some sense), then the symmetry center should maximise $D^{\rm loc}_P(\,\cdot\,)$; according to (P3), $D^{\rm loc}_P(\,\cdot\,)$ should be monotone non-increasing along any halfline originating from any P-deepest point; finally, (P4) states that as θ exits any compact set in $\mathbb{R}^k$, its depth should converge to zero. There is now an almost universal agreement in the literature that (P1)-(P4) are the natural desirable properties for location depths.

In view of this, one may wonder what are the desirable properties for a scatter depth. Inspired by (P1)-(P4), we argue that a generic scatter depth $D^{\rm sc}_P(\,\cdot\,) : \mathcal{P}_k \to [0,1]$ should satisfy the following properties, all involving an (unless otherwise specified) arbitrary probability measure P over $\mathbb{R}^k$:

(Q1) *Affine invariance*: for any $A \in GL_k$ and $b \in \mathbb{R}^k$, $D^{\rm sc}_{P_{A,b}}(A\Sigma A') = D^{\rm sc}_P(\Sigma)$, where $P_{A,b}$ is still as defined on page 16;

(Q2) *Fisher consistency under ellipticity*: if P is elliptically symmetric with location θ_0 and scatter Σ_0, then $D^{\rm sc}_P(\Sigma_0) \geq D^{\rm sc}_P(\Sigma)$ for any $\Sigma \in \mathcal{P}_k$;

(Q3) *Monotonicity relative to any deepest scatter*: if Σ_a maximises $D_P^{\rm sc}(\,\cdot\,)$, then, for any $\Sigma_b \in \mathcal{P}_k$, $t \mapsto D_P^{\rm sc}((1-t)\Sigma_a + t\Sigma_b)$ is monotone non-increasing over $[0,1]$;

(Q4) *Vanishing at the boundary of the parameter space*: if (Σ_n) F-converges to the boundary of $\mathcal{P}_k$ (in the sense that either $d_F(\Sigma_n, \Sigma) \to 0$ for some $\Sigma \in \mathcal{S}_k \setminus \mathcal{P}_k$ or $d_F(\Sigma_n, I_k) \to \infty$), then $D_P^{\rm sc}(\Sigma_n) \to 0$.

While (Q1) and (Q3) are the natural scatter counterparts of (P1) and (P3), respectively, some comments are in order for (Q2) and (Q4). We start with (Q2). In essence, (P2) requires that, whenever an indisputable location center exists (as it is the case for symmetric distributions), this location should be flagged as most central by the location depth at hand. A similar reasoning leads to (Q2): we argue that, for an elliptical probability measure, the "true" value of the scatter parameter is indisputable, and (Q2) then imposes that the scatter depth at hand should identify this true scatter value as the (or at least, as a) deepest one. One might actually strengthen (Q2) by replacing the elliptical model there by a broader model in which the true scatter would still be clearly defined. In such a case, of course, the larger the model for which scatter depth satisfies (Q2), the better (a possibility, that we do not explore here, is to consider the union of the elliptical model and the *independent component model*; see Ilmonen & Paindaveine, 2011 and the references therein). This is parallel to what happens in (P2): the weaker the symmetry assumption under which (P2) is satisfied, the better (for instance, having (P2) satisfied with angular symmetry is better than having it satisfied with central symmetry only); see Zuo & Serfling (2000).

We then turn to (Q4), whose location counterpart (P4) is typically read by saying that the depth/centrality $D_P^{\rm loc}(\theta_n)$ goes to zero when the point θ_n goes to the boundary of the *sample space*. In the spirit of parametric depth (Mizera, 2002; Mizera & Müller, 2004), however, it is more appropriate to look at θ_n as a candidate location fit and to consider that (P4) imposes that the appropriateness $D_P^{\rm loc}(\theta_n)$ of this fit goes to zero as θ_n goes to the boundary of the *parameter space*. For location, the confounding between the sample space and parameter space (both are $\mathbb{R}^k$) allows for both interpretations. For scatter, however, there is no such confounding (the sample space is $\mathbb{R}^k$ and the parameter space is $\mathcal{P}_k$), and we argue (Q4) above is the natural scatter version of (P4): whenever Σ_n goes to the boundary of the parameter space $\mathcal{P}_k$, scatter depth should flag it as an arbitrarily poor candidate fit.

Theorem 2.1 states that scatter halfspace depth satisfies (Q1) as soon as it is based on an affine-equivariant T. Scatter halfspace depth satisfies (Q3) as well: if Σ_a maximises $HD_{P,T}^{\rm sc}(\,\cdot\,)$, then Theorem 3.3 indeed readily implies that

$$HD_{P,T}^{\rm sc}((1-t)\Sigma_a + t\Sigma_b) \geq \min(HD_{P,T}^{\rm sc}(\Sigma_a), HD_{P,T}^{\rm sc}(\Sigma_b)) = HD_{P,T}^{\rm sc}(\Sigma_b)$$

for any $\Sigma_b \in \mathcal{P}_k$ and $t \in [0,1]$. The next Fisher consistency result shows that, provided that T is affine-equivariant, (Q2) is also met.

Theorem 5.1. Let P be an elliptical probability measure over $\mathbb{R}^k$ with location θ_0 and scatter Σ_0, and let T be an affine-equivariant location functional. Then, (i) $HD_{P,T}^{\rm sc}(\Sigma_0) \geq HD_{P,T}^{\rm sc}(\Sigma)$ for any $\Sigma \in \mathcal{P}_k$, and the equality holds if and only if ${\rm Sp}(\Sigma_0^{-1}\Sigma) \subset \mathcal{I}_{\rm MSD}[Z_1]$, where $Z = (Z_1, \ldots, Z_k)' \stackrel{\mathcal{D}}{=} \Sigma_0^{-1/2}(X - \theta_0)$; (ii) in particular, if $\mathcal{I}_{\rm MSD}[Z_1]$ is a singleton (equivalently, if $\mathcal{I}_{\rm MSD}[Z_1] = \{1\}$), then $\Sigma \mapsto HD_{P,T}^{\rm sc}(\Sigma)$ is uniquely maximised at Σ_0.

While (Q1)-(Q3) are satisfied by scatter halfspace depth without any assumption on P, (Q4) is not, as the mixture example considered in Section 3 shows (since the sequence (Σ_n) considered there has limiting depth $1-s>0$). However, Theorem 3.2 reveals that (Q4) may fail only when $d_F(\Sigma_n,\Sigma)\to 0$ for some $\Sigma\in\mathcal{S}_k\setminus\mathcal{P}_k$. More importantly, Theorem 4.2 implies that T-scatter halfspace depth will satisfy (Q4) at any P that is smooth at T_P.

In a generic parametric depth setup, (Q3) would require that the parameter space is convex. If the parameter space rather is a non-flat Riemannian manifold, then it is natural to replace the "linear" monotonicity property (Q3) with a "geodesic" one. In the context of scatter depth, this would lead to replacing (Q3) with

($\widetilde{\text{Q3}}$) *Geodesic monotonicity relative to any deepest scatter*: if Σ_a maximises $D^{\rm sc}_P(\,\cdot\,)$, then, for any $\Sigma_b\in\mathcal{P}_k$, $t\mapsto D^{\rm sc}_P(\tilde{\Sigma}_t)$ is monotone non-increasing over $[0,1]$ along the geodesic path $\tilde{\Sigma}_t$ from Σ_a to Σ_b in (4.2).

We refer to Section 7 for a parametric framework where (Q3) cannot be considered and where ($\widetilde{\text{Q3}}$) needs to be adopted instead. For scatter, however, the hybrid nature of $\mathcal{P}_k$, which is both flat (as a convex subset of the vector space $\mathcal{S}_k$) and curved (as a Riemannian manifold with non-positive curvature), allows to consider both (Q3) and ($\widetilde{\text{Q3}}$). Just like (Q3) follows from quasi-concavity of the mapping $\Sigma\mapsto HD^{\rm sc}_{P,T}(\Sigma)$, ($\widetilde{\text{Q3}}$) would follow from the same mapping being *geodesic quasi-concave*, in the sense that $HD_{P,T}(\tilde{\Sigma}_t)\geq\min(HD_{P,T}(\Sigma_a),HD_{P,T}(\Sigma_b))$ along the geodesic path $\tilde{\Sigma}_t$ from Σ_a to Σ_b. Geodesic quasi-concavity would actually imply that scatter halfspace depth regions are *geodesic convex*, in the sense that, for any $\Sigma_a,\Sigma_b\in R^{\rm sc}_{P,T}(\alpha)$, the geodesic from Σ_a to Σ_b is contained in $R^{\rm sc}_{P,T}(\alpha)$. We refer to Dümbgen & Tyler (2016) for an application of geodesic convex functions to inference on (high-dimensional) scatter matrices.

Theorem 3.3 shows that $\Sigma\mapsto HD^{\rm sc}_{P,T}(\Sigma)$ is quasi-concave for any P. A natural question is then whether or not this extends to geodesic quasi-concavity. The answer is positive at any k-variate elliptical probability measure and at the k-variate probability measure with independent Cauchy marginals.

Theorem 5.2. Let P be an elliptical probability measure over $\mathbb{R}^k$ or the k-variate probability measure with independent Cauchy marginals, and let T be an affine-equivariant location functional. Then, (i) $\Sigma\mapsto HD^{\rm sc}_{P,T}(\Sigma)$ is geodesic quasi-concave, so that (ii) $R^{\rm sc}_{P,T}(\alpha)$ is geodesic convex for any $\alpha\geq 0$.

We close this section with a numerical illustration of the quasi-concavity results in Theorems 3.3 and 5.2 and with an example showing that geodesic quasi-concavity may actually fail to hold. Figure 4 provides, for three bivariate probability measures P, the plots of $t\mapsto HD^{\rm sc}_P(\Sigma_t)$ and $t\mapsto HD^{\rm sc}_P(\tilde{\Sigma}_t)$, where $\Sigma_t=(1-t)\Sigma_a+t\Sigma_b$ is the linear path from $\Sigma_a=I_2$ to $\Sigma_b=\mathrm{diag}(0.001,20)$ and where $\tilde{\Sigma}_t=\Sigma_a^{1/2}(\Sigma_a^{-1/2}\Sigma_b\Sigma_a^{-1/2})^t\Sigma_a^{1/2}$ is the corresponding geodesic path. The three distributions considered are (i) the bivariate normal with location zero and scatter I_2, (ii) the bivariate distribution with independent Cauchy marginals, and (iii) the empirical distribution associated with a random sample of size $n=200$ from the bivariate mixture distribution $P=\frac{1}{2}P_1+\frac{1}{4}P_2+\frac{1}{4}P_3$, where P_1 is the standard normal, P_2 is the normal with mean $(0,4)'$ and covariance matrix $\frac{1}{10}I_2$, and P_3 is the normal with mean $(0, -4)'$ and covariance matrix $\frac{1}{10} I_2$. Figure 4 illustrates that (linear) quasi-concavity of scatter halfspace depth always holds, but that geodesic quasi-concavity may fail to hold. Despite this counterexample, extensive numerical experiments led us to think that geodesic quasi-concavity is the rule rather than the exception.

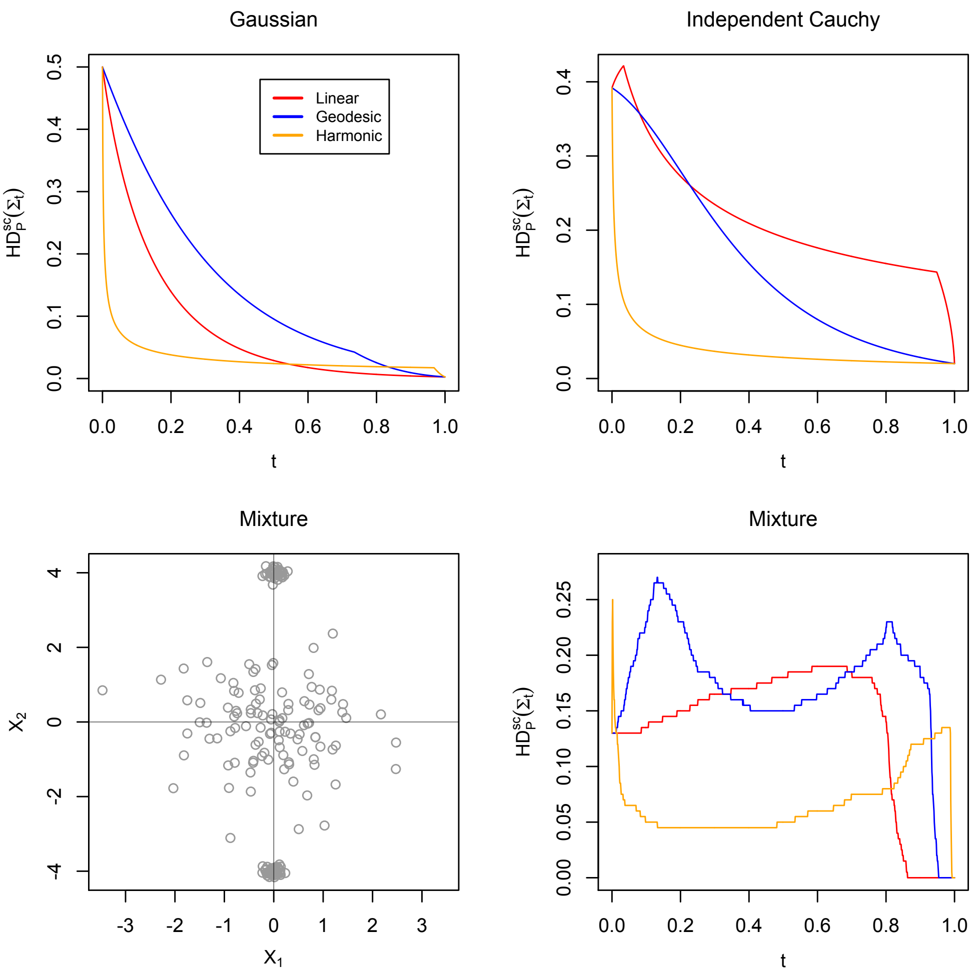

Figure 4: *Plots, for various bivariate probability measures P, of the scatter halfspace depth function $\Sigma \mapsto HD_P^{sc}(\Sigma)$ along the linear path $\Sigma_t = (1-t)\Sigma_a + t\Sigma_b$ (red), the geodesic path $\tilde{\Sigma}_t = \Sigma_a^{1/2}(\Sigma_a^{-1/2}\Sigma_b\Sigma_a^{-1/2})^t\Sigma_a^{1/2}$ (blue), and the harmonic path $\Sigma_t^* = ((1-t)\Sigma_a^{-1} + t\Sigma_b^{-1})^{-1}$ (orange), from $\Sigma_a = I_2$ to $\Sigma_b = \mathrm{diag}(0.001, 20)$; harmonic paths are introduced in Section 6. The probability measures considered are the bivariate normal with location zero and scatter I_2 (top left), the bivariate distribution with independent Cauchy marginals (top right), and the empirical probability measure associated with a random sample of size $n = 200$ from the bivariate mixture distribution described in Section 5 (bottom right). The scatter plot of the sample used in the mixture case is provided in the bottom left panel.*

6 Concentration halfspace depth

In various setups, the parameter of interest is the *concentration matrix* $\Gamma := \Sigma^{-1}$ rather than the scatter matrix Σ. For instance, in Gaussian graphical models, the (i, j)-entry of Γ is zero if and only if the ith and jth marginals are conditionally independent given all other marginals. It may then be useful to define a depth for inverse scatter matrices. The scatter halfspace depth in (2.1) naturally leads to defining the T-*concentration halfspace depth* of Γ with respect to P as

$$HD^{\rm conc}_{P,T}(\Gamma) := HD^{\rm sc}_{P,T}(\Gamma^{-1})$$

and the corresponding T-*concentration halfspace depth regions* as $R^{\rm conc}_{P,T}(\alpha) := \{\Gamma \in \mathcal{P}_k : HD^{\rm conc}_{P,T}(\Gamma) \geq \alpha\}$, $\alpha \geq 0$. As indicated by an anonymous referee, the definition of T-concentration halfspace depth alternatively results, through the use of "innovated transformation" (see, e.g., Hall & Jin, 2010, Fan *et al.*, 2013, or Fan & Lv, 2016), from the concept of (an affine-invariant) T-scatter halfspace depth.

Concentration halfspace depth and concentration halfspace depth regions inherit the properties of their scatter antecedents, sometimes with subtle modifications. The former is affine-invariant and the latter are affine-equivariant as soon as they are based on an affine-equivariant location functional T. Concentration halfspace depth is upper F- and g-semicontinuous for any probability measure P (so that the regions $R^{\rm conc}_{P,T}(\alpha)$ are F- and g-closed) and F- and g-continuous if P is smooth at T_P. While the regions $R^{\rm conc}_{P,T}(\alpha)$ are still g-bounded (hence also, F-bounded) for $\alpha > \alpha_{P,T}$, the outer regions $R^{\rm conc}_{P,T}(\alpha)$, $\alpha \leq \alpha_{P,T}$, here may fail to be F-bounded (this is because implosion of Σ, under which scatter halfspace depth may fail to go below $\alpha_{P,T}$, is associated with explosion of Σ^{-1}). Finally, uniform consistency and existence of a concentration halfspace deepest matrix are guaranteed under the same conditions on P and T as for scatter halfspace depth.

Quasi-concavity of concentration halfspace depth and convexity of the corresponding regions require more comments. The linear path $t \mapsto (1-t)\Gamma_a + t\Gamma_b$ between the concentration matrices $\Gamma_a = \Sigma_a^{-1}$ and $\Gamma_b = \Sigma_b^{-1}$ determines a *harmonic path* $t \mapsto \Sigma^*_t := ((1-t)\Sigma_a^{-1} + t\Sigma_b^{-1})^{-1}$ between the corresponding scatter matrices Σ_a and Σ_b. In line with the definitions adopted in the previous sections, we will say that $f : \mathcal{P}_k \to \mathbb{R}$ is *harmonic quasi-concave* if $f(\Sigma^*_t) \geq \min(f(\Sigma_a), f(\Sigma_b))$ for any $\Sigma_a, \Sigma_b \in \mathcal{P}_k$ and $t \in [0,1]$, and that a subset R of $\mathcal{P}_k$ is *harmonic convex* if $\Sigma_a, \Sigma_b \in R$ implies that $\Sigma^*_t \in R$ for any $t \in [0,1]$. Clearly, concentration halfspace depth is quasi-concave if and only if scatter halfspace depth is harmonic quasi-concave, which turns out to be the case in the elliptical and independent Cauchy cases. We thus have the following result.

Theorem 6.1. Let P be an elliptical probability measure over $\mathbb{R}^k$ or the k-variate probability measure with independent Cauchy marginals, and let T be an affine-equivariant location functional. Then, (i) $\Gamma \mapsto HD^{\rm conc}_{P,T}(\Gamma)$ is quasi-concave, so that (ii) $R^{\rm conc}_{P,T}(\alpha)$ is convex for any $\alpha \geq 0$.

However, concentration halfspace depth may fail to be quasi-concave, since, as we show by considering the mixture example in Figure 4, scatter halfspace depth may fail to be harmonic quasi-concave. The figure, that also plots scatter halfspace depth along harmonic paths, confirms that, while scatter halfspace depth is harmonic quasi-concave for the Gaussian and independent Cauchy examples there, it is not in the mixture example. In this mixture example, thus, concentration halfspace depth fails to be quasi-concave and the corresponding depth regions fail to be convex. This is not a problem per se — recall that famous (location) depth functions, like, e.g., the simplicial depth from Liu (1990), may provide non-convex depth regions.

For completeness, we present the following result which shows that some form of quasi-concavity for concentration halfspace depth survives.

Theorem 6.2. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, (i) $\Gamma \mapsto HD^{\mathrm{conc}}_{P,T}(\Gamma)$ is harmonic quasi-concave, so that (ii) $R^{\mathrm{conc}}_{P,T}(\alpha)$ is harmonic convex for any $\alpha \geq 0$.

Since concentration halfspace depth is harmonic quasi-concave if and only if scatter halfspace depth is quasi-concave, the result is a direct corollary of Theorem 3.3. Quasi-concavity and harmonic quasi-concavity clearly are dual concepts, relative to scatter and concentration halfspace depths (which justifies the * notation in the path Σ^*_t, dual to Σ_t). Interestingly, $\Gamma \mapsto HD^{\mathrm{conc}}_{P,T}(\Gamma)$ is geodesic quasi-concave if and only if $\Sigma \mapsto HD^{\mathrm{sc}}_{P,T}(\Sigma)$ is, so that concentration halfspace depth regions are geodesic convex if and only if scatter halfspace depth regions are.

7 Shape halfspace depth

In many multivariate statistics problems (PCA, CCA, sphericity testing, etc.), it is sufficient to know the scatter matrix Σ up to a positive scalar factor. In PCA, for instance, all scatter matrices of the form $c\Sigma$, $c > 0$, indeed provide the same unit eigenvectors $v_\ell(c\Sigma)$, $\ell = 1, \ldots, k$, hence the same principal components. Moreover, when it comes to deciding how many principal components to work with, a common practice is to look at the proportions of explained variance $\sum_{\ell=1}^m \lambda_\ell(c\Sigma)/\sum_{\ell=1}^k \lambda_\ell(c\Sigma)$, $m = 1, \ldots, k-1$, which do not depend on c either. In PCA, thus, the parameter of interest is a *shape matrix*, that is, a normald version, V say, of the scatter matrix Σ.

The generic way to normalise a scatter matrix Σ into a shape matrix V is based on a scale functional S, that is, on a mapping $S : \mathcal{P}_k \to \mathbb{R}^+_0$ satisfying (i) $S(I_k) = 1$ and (ii) $S(c\Sigma) = cS(\Sigma)$ for any $c > 0$ and $\Sigma \in \mathcal{P}_k$. In the following, we will further assume that (iii) if $\Sigma_1, \Sigma_2 \in \mathcal{P}_k$ satisfy $\Sigma_2 \geq \Sigma_1$ (in the sense that $\Sigma_2 - \Sigma_1$ is positive semidefinite), then $S(\Sigma_2) \geq S(\Sigma_1)$. Such a scale functional leads to factorising $\Sigma(\in \mathcal{P}_k)$ into $\Sigma = \sigma^2_S V_S$, where $\sigma^2_S := S(\Sigma)$ is the *scale* of Σ and $V_S := \Sigma/S(\Sigma)$ is its *shape matrix* (in the sequel, we will drop the subscript S in V_S to avoid overloading the notation). The resulting collection of shape matrices V will be denoted as $\mathcal{P}^S_k$. Note that the constraint $S(I_k) = 1$ ensures that, irrespective of the scale functional S adopted, I_k is a shape matrix. Common scale functionals satisfying (i)-(iii) are (a) $S_{\mathrm{tr}}(\Sigma) = (\mathrm{tr}\,\Sigma)/k$, (b) $S_{\det}(\Sigma) = (\det \Sigma)^{1/k}$, (c) $S^*_{\mathrm{tr}}(\Sigma) = k/(\mathrm{tr}\,\Sigma^{-1})$, and (d) $S_{11}(\Sigma) = \Sigma_{11}$; we refer to Paindaveine & Van Bever (2014) for references where the scale functionals (a)-(d) are used. The corresponding shape matrices V are then normalised in such a way that (a) $\mathrm{tr}[V] = k$, (b) $\det V = 1$, (c) $\mathrm{tr}[V^{-1}] = k$, or (d) $V_{11} = 1$.

In this section, we propose a concept of halfspace depth for shape matrices. More precisely, for a probability measure P over $\mathbb{R}^k$, we define the (S,T)-*shape halfspace depth* of $V(\in \mathcal{P}^S_k)$ with respect to P as

$$HD^{\mathrm{sh},S}_{P,T}(V) := \sup_{\sigma^2 > 0} HD^{\mathrm{sc}}_{P,T}(\sigma^2 V), \tag{7.1}$$

where $HD^{\mathrm{sc}}_{P,T}(\sigma^2 V)$ is the T-scatter halfspace depth of $\sigma^2 V$ with respect to P. The corresponding depth regions are defined as

$$R^{\mathrm{sh},S}_{P,T}(\alpha) := \{V \in \mathcal{P}^S_k : HD^{\mathrm{sh},S}_{P,T}(V) \geq \alpha\}$$

(alike scatter, we will drop the index T in $HD^{\mathrm{sh},S}_{P,T}(V)$ and $R^{\mathrm{sh},S}_{P,T}(\alpha)$ whenever T is the Tukey median). The halfspace deepest shape (if any) is obtained by maximising the "profile depth" in (7.1), in the same way a profile likelihood approach would be based on the maximisation of a (shape) profile likelihood of the form $L^{\mathrm{sh}}_V = \sup_{\sigma^2>0} L_{\sigma^2 V}$. To the best of our knowledge, such a profile depth construction has never been considered in the literature.

We start the study of shape halfspace depth by considering our running, Gaussian and independent Cauchy, examples. For the k-variate normal with location θ_0 and scatter Σ_0 (hence, with S-shape matrix $V_0 = \Sigma_0/S(\Sigma_0)$),

$$\sigma^2 \mapsto HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) = 2\min\left(\Phi\left(\frac{b\sigma\lambda_k^{1/2}(V_0^{-1}V)}{\sqrt{S(\Sigma_0)}}\right) - \frac{1}{2}, 1 - \Phi\left(\frac{b\sigma\lambda_1^{1/2}(V_0^{-1}V)}{\sqrt{S(\Sigma_0)}}\right)\right)$$

(see (2.6)) will be uniquely maximised at the σ^2-value for which both arguments of the minimum are equal. It follows that

$$HD^{\mathrm{sh},S}_{P,T}(V) = 2\Phi\big(c(V_0^{-1}V)\lambda_k^{1/2}(V_0^{-1}V)\big) - 1,$$

where $c(\Upsilon)$ is the unique solution of $\Phi\big(c(\Upsilon)\lambda_k^{1/2}(\Upsilon)\big) - \frac{1}{2} = 1 - \Phi\big(c(\Upsilon)\lambda_1^{1/2}(\Upsilon)\big)$. At the k-variate distribution with independent Cauchy marginals, we still have that (with the same notation as in (2.7))

$$HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) = 2\min\left(\Psi\big(\sigma/\max_s(s'V^{-1}s)^{1/2}\big) - \tfrac{1}{2}, 1 - \Psi\big(\sigma\sqrt{\max(\mathrm{diag}(V))}\big)\right)$$

is maximised for fixed V when both arguments of the minimum are equal, that is, when $\sigma^2 = \big(\max_s(s'V^{-1}s)/\max(\mathrm{diag}(V))\big)^{1/2}$. Therefore,

$$\begin{aligned} HD^{\mathrm{sh},S}_{P,T}(V) &= 2\,\Psi\Big(\big(\max_s(s'V^{-1}s)\max(\mathrm{diag}(V))\big)^{-1/4}\Big) - 1 \\ &= \frac{2}{\pi}\arctan\Big(\big(\max_s(s'V^{-1}s)\max(\mathrm{diag}(V))\big)^{-1/4}\Big). \end{aligned}$$

Figure 5 draws, for six probability measures P and any affine-equivariant T, contour plots of $(V_{11}, V_{12}) \mapsto HD^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(V)$, where $HD^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(V)$ is the shape halfspace depth of $V = \left(\begin{smallmatrix} V_{11} & V_{12} \\ V_{12} & 2-V_{11}\end{smallmatrix}\right)$ with respect to P. Letting $\Sigma_A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1\end{smallmatrix}\right)$, $\Sigma_B = \left(\begin{smallmatrix} 4 & 0 \\ 0 & 1\end{smallmatrix}\right)$ and $\Sigma_C = \left(\begin{smallmatrix} 3 & 1 \\ 1 & 1\end{smallmatrix}\right)$, the probability measures P considered are those associated (i) with the bivariate normal distributions with location zero and scatter Σ_A, Σ_B and Σ_C, and (ii) with the distributions of $\Sigma_A^{1/2}Z$, $\Sigma_B^{1/2}Z$ and $\Sigma_C^{1/2}Z$, where Z has independent Cauchy marginals. Note that the maximal depth is larger in the Gaussian cases than in the Cauchy ones, that depth monotonically decreases along any ray originating from the deepest shape matrix and that it goes to zero if and only if the shape matrix converges to the boundary of the parameter space. Shape halfspace depth contours are smooth in the Gaussian cases but not in the Cauchy ones.

In both the Gaussian and independent Cauchy examples above, the supremum in (7.1) is a maximum. For empirical probability measures P_n, this trivially is always the case since $HD^{\rm sc}_{P_n,T}(\sigma^2 V)$ then takes its values in $\{\ell/n : \ell = 0, 1, \ldots, n\}$. The following result implies in particular that a sufficient condition for this supremum to be a maximum is that P is smooth at T_P (which is the case in both our running examples above).

Theorem 7.1. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Fix $V \in \mathcal{P}_k^S$ such that $cV \in R^{\rm sc}_{P,T}(\alpha_{P,T})$ for some $c > 0$. Then, $HD^{{\rm sh},S}_{P,T}(V) = HD^{\rm sc}_{P,T}(\sigma_V^2 V)$ for some $\sigma_V^2 > 0$.

The following affine-invariance/equivariance and uniform consistency results are easily obtained from their scatter antecedents.

Theorem 7.2. Let T be an affine-equivariant location functional. Then, (i) shape halfspace depth is affine-invariant in the sense that, for any probability measure P over $\mathbb{R}^k$, $V \in \mathcal{P}_k^S$, $A \in GL_k$ and $b \in \mathbb{R}^k$, we have $HD^{{\rm sh},S}_{P_{A,b},T}(AVA'/S(AVA')) = HD^{{\rm sh},S}_{P,T}(V)$, where $P_{A,b}$ is as defined on page 16. Consequently, (ii) shape halfspace depth regions are affine-equivariant, in the sense that $R^{{\rm sh},S}_{P_{A,b},T}(\alpha) = \big\{AVA'/S(AVA') : V \in R^{{\rm sh},S}_{P,T}(\alpha)\big\}$ for any probability measure P over $\mathbb{R}^k$, $\alpha \geq 0$, $A \in GL_k$ and $b \in \mathbb{R}^k$.

Theorem 7.3. Let P be a smooth probability measure over $\mathbb{R}^k$ and T be a location functional. Let P_n denote the empirical probability measure associated with a random sample of size n from P and assume that $T_{P_n} \to T_P$ almost surely as $n \to \infty$. Then $\sup_{V \in \mathcal{P}_k^S} |HD^{{\rm sh},S}_{P_n,T}(V) - HD^{{\rm sh},S}_{P,T}(V)| \to 0$ almost surely as $n \to \infty$.

Shape halfspace depth inherits the F- and g-continuity properties of scatter halfspace depth (Theorems 3.1 and 4.1, respectively), at least for a smooth P. More precisely, we have the following result.

Theorem 7.4. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, (i) $V \mapsto HD^{{\rm sh},S}_{P,T}(V)$ is upper F- and g-semicontinuous on $R^{{\rm sh},S}_{P,T}(\alpha_{P,T})$, so that (ii) for any $\alpha \geq \alpha_{P,T}$, the depth region $R^{{\rm sh},S}_{P,T}(\alpha)$ is F- and g-closed. (iii) If P is smooth at T_P, then $V \mapsto HD^{{\rm sh},S}_{P,T}(V)$ is F- and g-continuous.

The g-boundedness part of the following result will play a key role when proving the existence of a halfspace deepest shape.

Theorem 7.5. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Then, for any $\alpha > \alpha_{P,T}$, $R^{{\rm sh},S}_{P,T}(\alpha)$ is F- and g-bounded, hence g-compact. If $s_{P,T} \geq 1/2$, then this result is trivial in the sense that $R^{{\rm sh},S}_{P,T}(\alpha)$ is empty for $\alpha > \alpha_{P,T}$.

Comparing with the scatter result in Theorem 3.2, the shape result for F-boundedness requires the additional condition $\alpha > \alpha_{P,T}$ (for g-boundedness, this condition was already required in Theorem 4.2). This condition is actually necessary for scale functionals S for which implosion of a shape matrix V cannot be obtained without explosion, as it is the case, e.g., for $S_{\rm det}$ (the product of the eigenvalues of an $S_{\rm det}$-shape matrix being equal to one, the smallest eigenvalue of V cannot go to zero without the largest going to infinity). We illustrate this on the bivariate discrete example discussed below Theorem 4.2, still with an

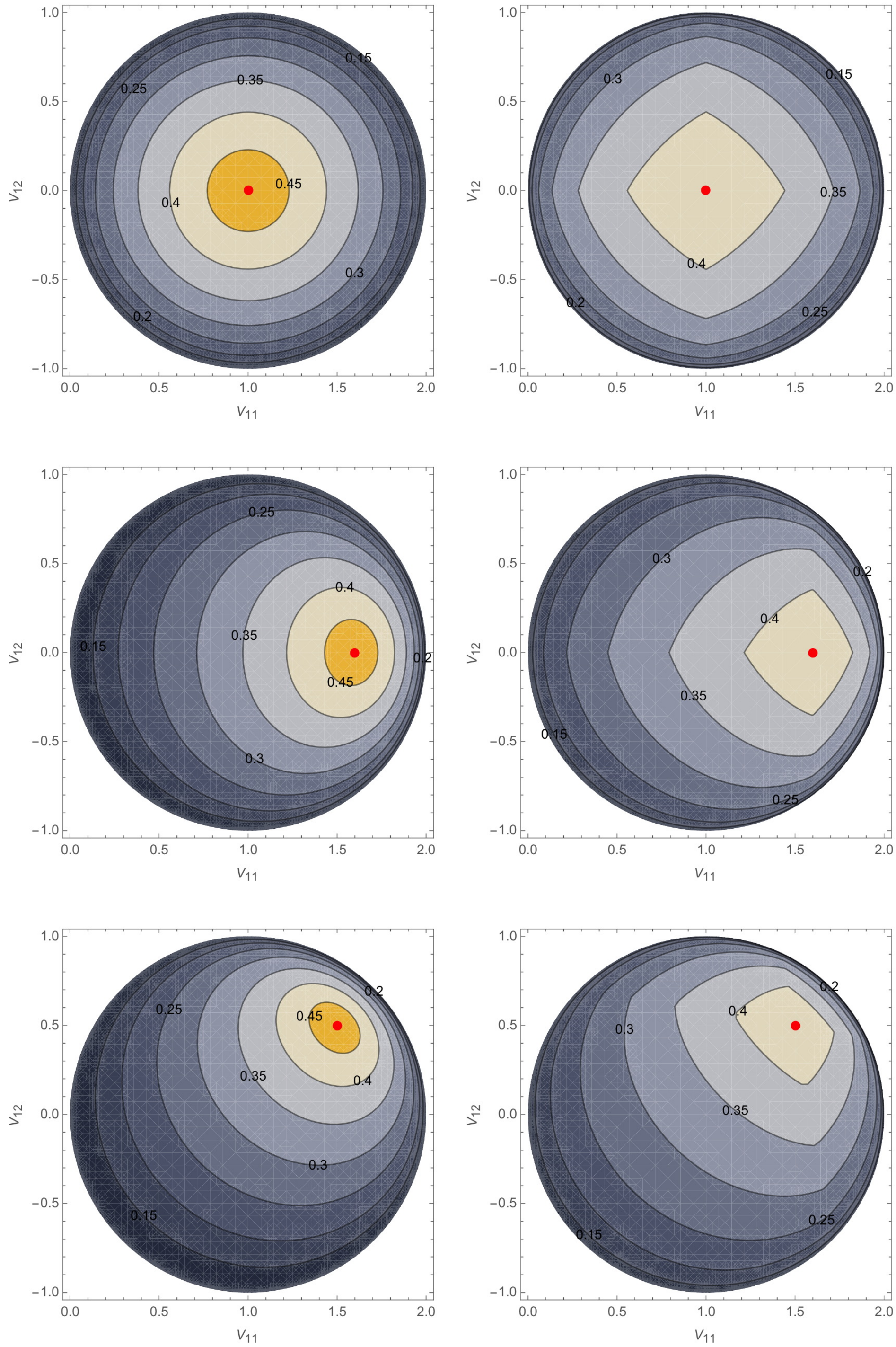

Figure 5: *Contour plots of* $(V_{11}, V_{12}) \mapsto HD^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(V)$*, for several bivariate probability measures* P *and an arbitrary affine-equivariant location functional* T*, where* $HD^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(V)$ *is the shape halfspace depth, with respect to* P*, of* $V = \begin{pmatrix} V_{11} & V_{12} \\ V_{12} & 2-V_{11} \end{pmatrix}$*. Letting* $\Sigma_A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$*,* $\Sigma_B = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$ *and* $\Sigma_C = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}$*, the probability measures* P *considered are those associated (i) with the bivariate normal distributions with location zero and scatter* Σ_A*,* Σ_B *and* Σ_C *(top, middle and bottom left), and (ii) with the distributions of* $\Sigma_A^{1/2}Z$*,* $\Sigma_B^{1/2}Z$ *and* $\Sigma_C^{1/2}Z$*, where* Z *has mutually independent Cauchy marginals (top, middle and bottom right). In each case, the "true"* S_{tr}*-shape matrix is marked in red.*

arbitrary centro-equivariant T. The sequence of scatter matrices $\Sigma_n = \mathrm{diag}(\frac{1}{n}, 1)$ there defines a sequence of S_{det}-shape matrices $V_n = \mathrm{diag}(\frac{1}{\sqrt{n}}, \sqrt{n})$, that is neither F- nor g-bounded. Since $HD^{\mathrm{sh},S_{\mathrm{det}}}_{P,T}(V_n) \geq HD^{\mathrm{sc}}_{P,T}(\Sigma_n) \geq 1/3 = \alpha_{P,T}$ for any n, we conclude that $R^{\mathrm{sh},S_{\mathrm{det}}}_{P,T}(\alpha_{P,T})$ is both F- and g-unbounded. Note also that F-boundedness of $R^{\mathrm{sh},S}_{P,T}(\alpha)$ depends on S. In particular, it is easy to check that the condition $\alpha > \alpha_{P,T}$ for F-boundedness is not needed for the scale functional S^*_{tr} (that is, $R^{\mathrm{sh},S^*_{\mathrm{tr}}}_{P,T}(\alpha)$ is F-bounded for any $\alpha > 0$). Finally, one trivially has that all $R^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(\alpha)$'s are F-bounded since the corresponding collection of shape matrices, $\mathcal{P}^{S_{\mathrm{tr}}}_k$, itself is F-bounded. Unlike F-boundedness, g-boundedness results are homogeneous in S, which further suggests that the g-topology is the most appropriate one to study scatter/shape depths.

As announced, the g-part of Theorem 7.5 allows to show that a halfspace deepest shape exists under mild conditions. More precisely, we have the following result.

Theorem 7.6. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Assume that $R^{\mathrm{sh},S}_{P,T}(\alpha_{P,T})$ is non-empty. Then, $\alpha^S_{*P,T} := \sup_{V \in \mathcal{P}^S_k} HD^{\mathrm{sh},S}_{P,T}(V) = HD^{\mathrm{sh},S}_{P,T}(V_*)$ for some $V_* \in \mathcal{P}^S_k$.

Alike scatter, a sufficient condition for the existence of a halfspace deepest shape is thus that P is smooth at T_P. In particular, a halfspace deepest shape exists in the Gaussian and independent Cauchy examples. In the k-variate independent Cauchy case, it readily follows from Theorem 4.4 that, irrespective of the centro-equivariant T used, $HD^{\mathrm{sh},S}_{P,T}(V)$ is uniquely maximised at $V_* = I_k$, with corresponding maximal depth $\frac{2}{\pi}\arctan\left(k^{-1/4}\right)$. The next Fisher-consistency result states that, in the elliptical case, the halfspace deepest shape coincides with the "true" shape matrix.

Theorem 7.7. Let P be an elliptical probability measure over $\mathbb{R}^k$ with location θ_0 and scatter Σ_0, hence with S-shape matrix $V_0 = \Sigma_0/S(\Sigma_0)$, and let T be an affine-equivariant location functional. Then, (i) $HD^{\mathrm{sh},S}_{P,T}(V_0) \geq HD^{\mathrm{sh},S}_{P,T}(V)$ for any $V \in \mathcal{P}^S_k$; (ii) if $\mathcal{I}_{\mathrm{MSD}}[Z_1]$ is a singleton (equivalently, if $\mathcal{I}_{\mathrm{MSD}}[Z_1] = \{1\}$), where $Z = (Z_1, \ldots, Z_k)' \stackrel{\mathcal{D}}{=} \Sigma_0^{-1/2}(X - \theta_0)$, then $V \mapsto HD^{\mathrm{sh},S}_{P,T}(V)$ is uniquely maximised at V_0.

We conclude this section by considering quasi-concavity properties of shape halfspace depth and convexity properties of the corresponding depth regions. It should be noted that, for some scale functionals S, the collection $\mathcal{P}^S_k$ of S-shape matrices is not convex; for instance, neither $\mathcal{P}^{S_{\mathrm{det}}}_k$ nor $\mathcal{P}^{S^*_{\mathrm{tr}}}_k$ is convex, so that it does not make sense to investigate whether or not $V \mapsto HD^{\mathrm{sh},S}_{P,T}(V)$ is quasi-concave for these scale functionals. It does, however, for S_{tr} and S_{11}, and we have the following result.

Theorem 7.8. Let P be a probability measure over $\mathbb{R}^k$ and T be a location functional. Fix $S = S_{\mathrm{tr}}$ or $S = S_{11}$. Then, (i) $V \mapsto HD^{\mathrm{sh},S}_{P,T}(V)$ is quasi-concave, that is, for any $V_a, V_b \in \mathcal{P}^S_k$ and $t \in [0,1]$, $HD^{\mathrm{sh},S}_{P,T}(V_t) \geq \min(HD^{\mathrm{sh},S}_{P,T}(V_a), HD^{\mathrm{sh},S}_{P,T}(V_b))$, where we let $V_t := (1-t)V_a + tV_b$; (ii) for any $\alpha \geq 0$, $R^{\mathrm{sh},S}_{P,T}(\alpha)$ is convex.

As mentioned above, neither $\mathcal{P}^{S_{\mathrm{det}}}_k$ nor $\mathcal{P}^{S^*_{\mathrm{tr}}}_k$ are convex in the usual sense (unlike for S_{tr} and S_{11}, thus, a unique halfspace deepest shape could not be defined through barycenters but would rather require a center-of-mass approach as in (4.3)). However, $\mathcal{P}_k^{S_{\mathrm{det}}}$ is geodesic convex, which justifies studying the possible geodesic convexity of $R^{\mathrm{sh}}_{P,S_{\mathrm{det}}}(\alpha)$ (this provides a parametric framework for which the shape version of (Q3) in Section 5 cannot be considered and for which it is needed to adopt the corresponding Property ($\widetilde{\mathrm{Q3}}$) instead). Similarly, $\mathcal{P}_k^{S^*_{\mathrm{tr}}}$ is harmonic convex, so that it makes sense to investigate the harmonic convexity of $R^{\mathrm{sh},S^*_{\mathrm{tr}}}_{P,T}(\alpha)$. We have the following results.

Theorem 7.9. Let T be an affine-equivariant location functional and P be an arbitrary probability measure over $\mathbb{R}^k$ for which T-scatter halfspace depth is geodesic quasi-concave. Then, (i) $V \mapsto HD^{\mathrm{sh},S_{\mathrm{det}}}_{P,T}(V)$ is geodesic quasi-concave, so that (ii) $R^{\mathrm{sh},S_{\mathrm{det}}}_{P,T}(\alpha)$ is geodesic convex for any $\alpha \geq 0$.

Theorem 7.10. Let T be an affine-equivariant location functional and P be an arbitrary probability measure over $\mathbb{R}^k$ for which T-scatter halfspace depth is harmonic quasi-concave. Then, (i) $V \mapsto HD^{\mathrm{sh},S^*_{\mathrm{tr}}}_{P,T}(V)$ is harmonic quasi-concave, so that (ii) $R^{\mathrm{sh},S^*_{\mathrm{tr}}}_{P,T}(\alpha)$ is harmonic convex for any $\alpha \geq 0$.

Figure 6 draws, for an arbitrary affine-equivariant location functional T, contour plots of $(V_{11}, V_{12}) \mapsto HD^{\mathrm{sh},S}_{P,T}(V_S)$, for the scale functionals S_{tr}, S_{det} and S^*_{tr}, where $HD^{\mathrm{sh},S}_{P,T}(V_S)$ is the shape halfspace depth, with respect to P, of the S-shape $V_S (\in \mathcal{P}_2^S)$ with upper-left entry V_{11} and upper-right entry V_{12}. The probability measures P considered are those associated (i) with the bivariate normal distribution with location zero and scatter $\Sigma_C = \left(\begin{smallmatrix} 3 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, and (ii) with the distribution of $\Sigma_C^{1/2} Z$, where Z has mutually independent Cauchy marginals. For S_{tr}, S_{det} and S^*_{tr}, the figure also shows the linear, geodesic and harmonic paths, respectively, linking the ("true") S-shape associated with Σ_C and those associated with $\Sigma_A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $\Sigma_B = \left(\begin{smallmatrix} 4 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. The results illustrate the convexity of the regions $R^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(\alpha)$, along with the geodesic (resp., harmonic) convexity of the regions $R^{\mathrm{sh},S_{\mathrm{det}}}_{P,T}(\alpha)$ (resp., $R^{\mathrm{sh},S^*_{\mathrm{tr}}}_{P,T}(\alpha)$).

8 A real-data application

We close this chapter with a real-data application. In this section, we analyze the returns of the Nasdaq Composite and S&P500 indices from February 1st, 2015 to February 1st, 2017. During that period, for each trading day and for each index, we collected returns every 5 minutes (that is, the difference between the index at a given time and 5 minutes earlier, when available), resulting in usually 78 bivariate observations per day. Due to some missing values, the exact number of returns per day varies, and only days with at least 70 observations were considered. The resulting dataset comprises a total of 38489 bivariate returns distributed over $D = 478$ trading days.

The goal of this analysis is to determine which days, during the two-year period, exhibit an atypical behavior. In line with the fact that the main focus in finance is on volatily, atypicality here will refer to deviations from the "global" behavior either in scatter (i.e., returns do not follow the global dispersion pattern) or in scale only (i.e., returns show a usual shape but their overall size is different). Atypical days will be detected by comparing intraday estimates of scatter and shape with a global version.

Below, $\hat{\Sigma}_{\mathrm{full}}$ will denote the minimum covariance determinant (MCD) scatter estimate on

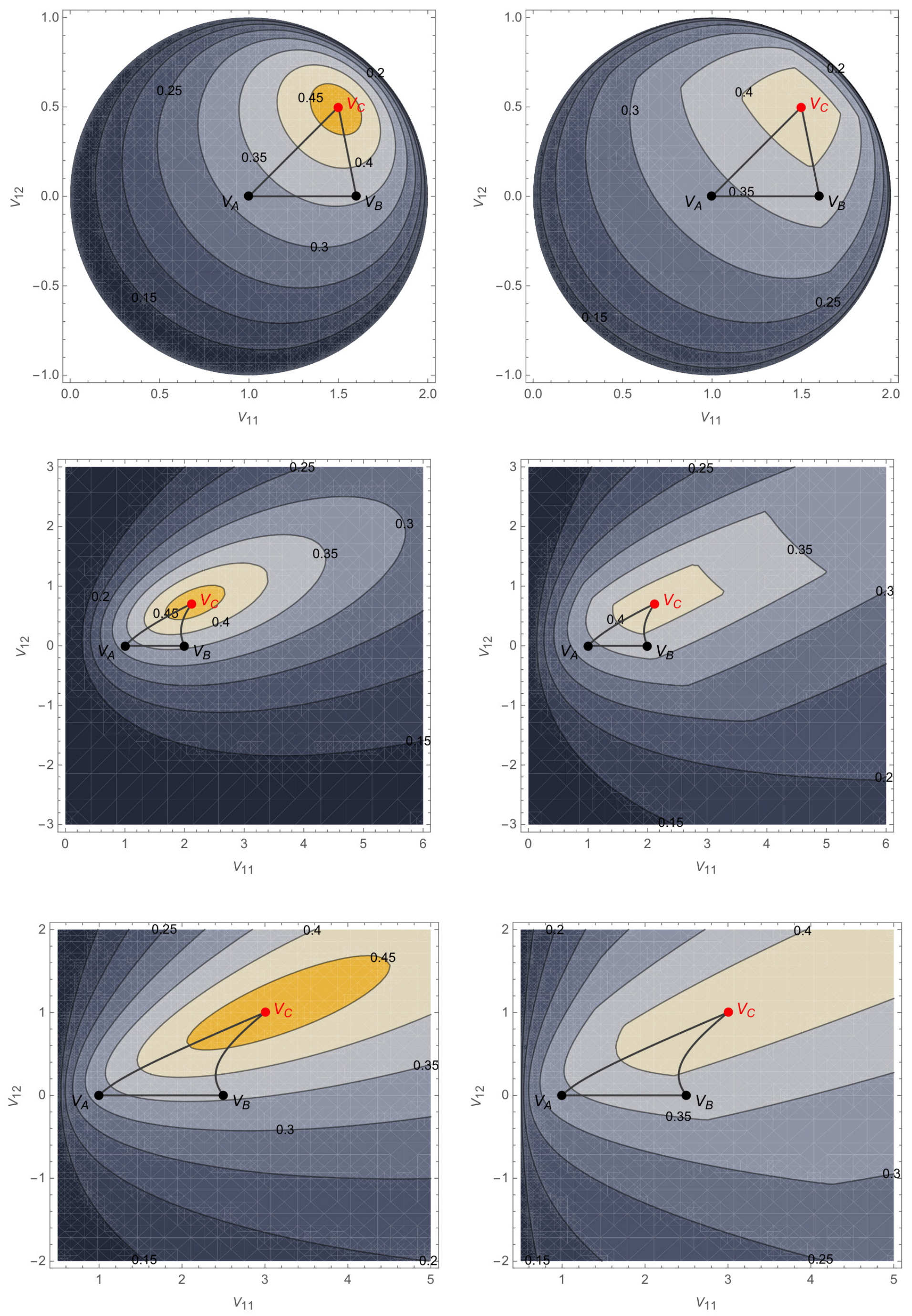

Figure 6: *Contour plots of $(V_{11}, V_{12}) \mapsto HD^{\mathrm{sh},S}_{P,T}(V_S)$, for S_{tr} (top), S_{det} (middle) and S^*_{tr} (bottom), where $HD^{\mathrm{sh},S}_{P,T}(V_S)$ is the shape halfspace depth, with respect to P, of the S-shape $V_S (\in \mathcal{P}_2^S)$ with upper-left entry V_{11} and upper-right entry V_{12}, for an arbitrary affine-equivariant location functional T. The probability measures P considered are those associated (i) with the bivariate normal distribution with location zero and scatter $\Sigma_C = \left(\begin{smallmatrix} 3 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ (left) and (ii) with the distribution of $\Sigma_C^{1/2} Z$, where Z has mutually independent Cauchy marginals (right). In each case, the ("true") S-shape associated with Σ_C is marked in red and those associated with Σ_A and Σ_B from Figure 5 are marked in black. Linear paths (top), geodesic paths (middle) and harmonic paths (bottom) between these three shapes are drawn.*

the empirical distribution P_{full} of the returns over the two-year period, and $\hat{V}_{\text{full}}$ will stand for the resulting shape estimate $\hat{V}_{\text{full}} = \hat{\Sigma}_{\text{full}}/S_{\text{det}}(\hat{\Sigma}_{\text{full}})$. For any $d = 1, \ldots, D$, $\hat{\Sigma}_d$ and $\hat{V}_d$ will denote the corresponding estimates on the empirical distribution P_d on day d.

The rationale behind the choice of MCD rather than standard covariance as an estimation method for scatter/shape is twofold. First, the former will naturally deal with outliers inherently arising in the data (the first few returns after an overnight or weekend break are famously more volatile and their importance should be downweighted in the estimation procedure). Second, as hinted above, the global estimate will provide a baseline to measure the atypicality of any given day, which will be done, among others, using its intraday depth. It would be natural to use halfspace deepest scatter/shape matrices on P_{full} as global estimates for scatter/shape. While locating the exact maxima is a non-trivial task, the MCD shape estimator has already a high depth value ($HD^{\text{sh},S_{\text{det}}}_{P_{\text{full}}}(\hat{V}_{\text{full}}) = 0.481$), which makes it a very good proxy for the halfspace deepest shape. For the same reason, the scaled MCD estimator $\bar{\Sigma}_{\text{full}} = \sigma^2_{\text{full}} V_{\text{full}}$ with $\sigma^2_{\text{full}} = \text{argmax}_{\sigma^2} HD^{\text{sc}}_{P_{\text{full}}}(\sigma^2 V_{\text{full}})$ (that, obviously, satisfies $HD^{\text{sc}}_{P_{\text{full}}}(\bar{\Sigma}_{\text{full}}) = 0.481$) is similarly an excellent proxy for the halfspace deepest scatter. In contrast, the shape estimate associated with the standard covariance matrix (resp., the deepest scaled version of the covariance matrix) has a global shape (resp., scatter) depth of only 0.426.

For each day, the following measures of (a)typicality (three for scatter, three for shape) are computed: (i) the scatter depth $HD^{\text{sc}}_{P_d}(\bar{\Sigma}_{\text{full}})$ of $\bar{\Sigma}_{\text{full}}$ in day d, (ii) the shape depth $HD^{\text{sh},S_{\text{det}}}_{P_d}(\hat{V}_{\text{full}})$ of $\hat{V}_{\text{full}}$ in day d, (iii) the scatter Frobenius distance $d_F(\hat{\Sigma}_d, \hat{\Sigma}_{\text{full}})$, (iv) the shape Frobenius distance $d_F(\hat{V}_d, \hat{V}_{\text{full}})$, (v) the scatter geodesic distance $d_g(\hat{\Sigma}_d, \hat{\Sigma}_{\text{full}})$, and (vi) the shape geodesic distance $d_g(\hat{V}_d, \hat{V}_{\text{full}})$. Of course, low depths or high distances point to atypical days. Practitioners might be tempted to base the distances in (iii)-(vi) on standard covariance estimates, which would actually provide poorer performances in the present outlier detection exercise (due to the masking effect resulting from using a non-robust global dispersion measure as a baseline). Here, we rather use MCD-based estimates to ensure a fair comparison with the depth-based methods in (i)-(ii).

Figure 7 provides the plots of the quantities in (i)-(vi) above as a function of d, $d = 1, \ldots, D$. Major events affecting the returns during the two years are marked there. They are (1) the Black Monday on August 24th, 2015 (orange) when world stock markets went down substantially, (2) the crude oil crisis on January 20th, 2016 (dark blue) when oil barrel prices fell sharply, (3) the Brexit vote aftermath on June 24th, 2016 (dark green), (4) the end of the low volatility period on September 13th, 2016 (red), (5) the Donald Trump election on November 9th, 2016 (purple), and (6) the announcement and aftermath of the federal rate hikes on December 14th, 2016 (teal).

Detecting atypical events was achieved by flagging outliers in either collections of scatter or shape depth values. This was conducted by constructing box-and-whisker plots of those collections and marking events with depth value below 1.5 IQR of the first quartile. This procedure flagged events (1), (2) and (6) as outlying in scatter and 21 days — including events (1), (2), (3) and (5) — as atypical in shape. Most of the resulting 22 outlying days can be associated (that is, are temporally close) to one of the events (1)-(6) above. For example, 9 days are flagged within the period extending from January 20th, 2016 to February 9th, 2016, during which continuous slump in oil prices rocked the marked strongly, with biggest loss for S&P 500 index on February 9th. Remarkably, out of the 22 flagged outliers, only two (namely October 1st, 2015 and December 14th, 2015) could not be associated with major events. Event (4), although not deemed outlying, was added to mark the end of the low volatility period.

Events (1) and (2) are noticeably singled out by all outlyingness measures, displaying low depth values and high Frobenius and geodesic distances, but the four remaining events tell a very different story. In particular, event (6) exhibits a low scatter depth but a relatively high shape depth, which means that this day shows a shape pattern that is in line with the global one but is very atypical in scale (that is, in volatility size). Quite remarkably, the four distances considered fail to flag this day as an atypical one. A similar behavior appears throughout the two-month period spanning July, August and early September 2016 (between events (3) and (4)), during which the markets have seen a historical streak of small volatility. This period presents widely varying scatter depth values together with stable and high shape depth values, which is perfectly in line with what has been seen on the markets, where only the volatility of the indices was low in days that were otherwise typical. Again, the four distance plots are blind to this relative behavior of scatter and shape in the period.

Events (3) to (5) are picked up by depth measures and scatter distances, though more markedly by the former. This is particularly so for event (3), which sticks out sharply in both depths. The fact that the scatter depth is even lower than the shape depth suggests that event (3) is atypical not only in shape but also in scale. Interestingly, distance measures fully miss the shape outlyingness of this event. Actually, shape distances do not assign large values to any of the events (3) to (6) and, from March 2016 onwards, these distances stay in the same range — particularly so for the Frobenius ones in (iv). In contrast, the better ability of shape depth to spot outlyingness may be of particular importance in cases where one wants to discard the overall volatility size to rather focus on the shape structure of the returns.

To summarise, the detection of atypical patterns in the dispersion of intraday returns can more efficiently be performed with scatter/shape depths than on the basis of distance measures. Arguably, the fact that the proposed depths use all observations and not a sole estimate of scatter/shape allows to detect deviations from global behaviors more sharply. As showed above, comparing scatter and shape depth values provides a tool that permits the distinction between shape and scale outliers.

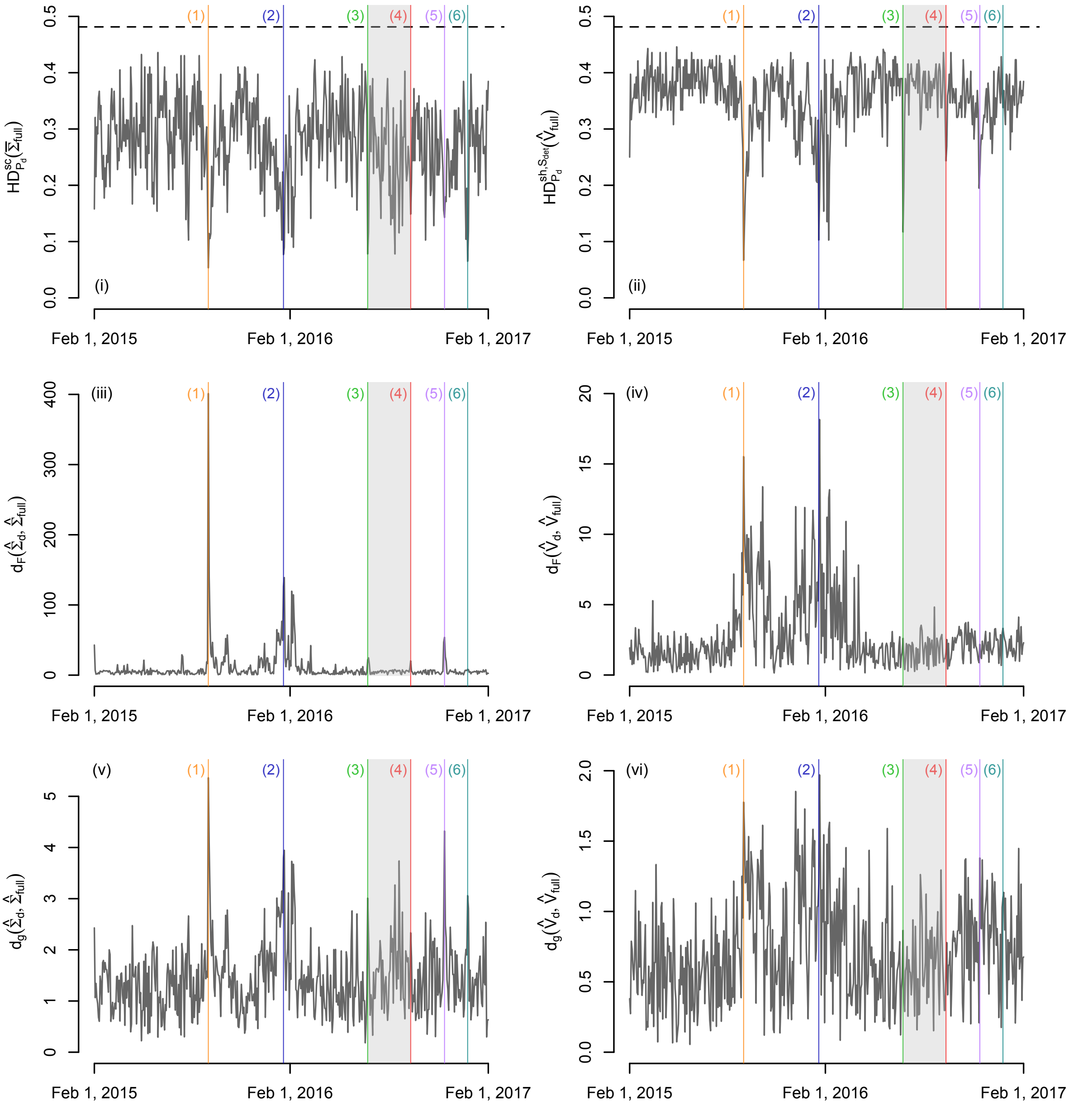

Figure 7: *Plots of (i)* $HD^{\mathrm{sc}}_{P_d}(\bar{\Sigma}_{\mathrm{full}})$*, (ii)* $HD^{\mathrm{sh},S_{\mathrm{det}}}_{P_d}(\hat{V}_{\mathrm{full}})$*, (iii)* $d_F(\hat{\Sigma}_d, \hat{\Sigma}_{\mathrm{full}})$*, (iv)* $d_F(\hat{V}_d, \hat{V}_{\mathrm{full}})$*, (v)* $d_g(\hat{\Sigma}_d, \hat{\Sigma}_{\mathrm{full}})$ *and (vi)* $d_g(\hat{V}_d, \hat{V}_{\mathrm{full}})$*, as a function of* d*, for the MCD scatter and shape estimates described in Section 8. The horizontal dotted lines in (i)-(ii) correspond to the global depths* $HD^{\mathrm{sc}}_{P_{\mathrm{full}}}(\bar{\Sigma}_{\mathrm{full}})$ *and* $HD^{\mathrm{sh},S_{\mathrm{det}}}_{P_{\mathrm{full}}}(\hat{V}_{\mathrm{full}})$*, respectively. All depths make use of the Tukey median as a location functional. Vertical lines mark the six events listed in Section 8.*

Chapter II

Tyler Shape Depth

Tyler Shape Depth[1]

Abstract In many problems from multivariate analysis (principal component analysis, testing for sphericity, etc.), the parameter of interest is a shape matrix, that is, a normalised version of the corresponding scatter or dispersion matrix. In this chapter, we propose a depth concept for shape matrices which is of a sign nature, in the sense that it involves data points only through their directions from the center of the distribution. We use the terminology Tyler shape depth since the resulting estimator of shape — namely, the deepest shape matrix — is the depth-based counterpart of the celebrated M-estimator of shape from Tyler (1987). We investigate the invariance, quasi-concavity and continuity properties of Tyler shape depth, as well as the topological and boundedness properties of the corresponding depth regions. We study existence of a deepest shape matrix and prove Fisher consistency in the elliptical case. We derive a Glivenko-Cantelli-type result and establish the almost sure consistency of the deepest shape matrix estimator. We also consider depth-based tests for shape and investigate their finite-sample performances through simulations. Finally, we illustrate the practical relevance of the proposed depth concept on a real data example.

Keywords: Elliptical distribution; Robustness; Shape matrix; Statistical depth; Test for sphericity.

[1]This chapter is a joint work with Davy Paindaveine. The manuscript has been (again, in a slightly different form) conditionally accepted for publication in *Biometrika*.

1 Introduction

Location depths measure the centrality of an arbitrary k-vector θ with respect to a probability measure $P = P^X$ over $\mathbb{R}^k$. Letting $\mathcal{S}^{k-1} = \{x \in \mathbb{R}^k : \|x\|^2 = x'x = 1\}$ be the unit sphere in $\mathbb{R}^k$, the most famous instance is the (Tukey, 1975) *halfspace depth*[2]

$$D(\theta, P) = \inf_{u \in \mathcal{S}^{k-1}} P[u'(X - \theta) \geq 0], \tag{1.1}$$

the lower bound of the probability mass of any halfspace whose boundary hyperplane contains θ. The halfspace depth regions $\{\theta \in \mathbb{R}^k : D(\theta, P) \geq \alpha\}$ form a family of nested convex subsets of $\mathbb{R}^k$. The innermost region $M_P = \{\theta \in \mathbb{R}^k : D(\theta, P) = \max_{\xi \in \mathbb{R}^k} D(\xi, P)\}$ (the maximum always exists; see Rousseeuw & Ruts, 1999) extends the univariate median interval to the multivariate case. Whenever a unique representative of M_P is needed, one often considers the *Tukey median* θ_P defined as the barycentre of M_P (from convexity, θ_P has maximal depth). The use of θ_P as a robust alternative to the expectation $E(X)$ is only one out of the numerous applications of halfspace depth. Many inference problems can indeed be tackled in a robust and nonparametric way by using the center-outward order resulting from depth; see, e.g., Liu *et al.* (1999). Halfspace depth is also important through its links with multivariate quantiles; see, e.g., Hallin *et al.* (2010) and the references therein.

Like in the previous chapter, the focus here is not on location parameters as above, but rather on some specific multivariate dispersion parameters, namely on *shape matrices*. We now describe shape in a context where its definition makes a large consensus, that is, in the elliptical setup. Denoting as $\mathcal{P}_k$ the collection of $k \times k$ symmetric positive definite matrices and writing $A^{1/2}$, with $A \in \mathcal{P}_k$, for the unique square root of A in $\mathcal{P}_k$, we will say that $P = P^X$ is elliptical with location $\theta (\in \mathbb{R}^k)$, *shape* $V (\in \mathcal{P}_{k,\mathrm{tr}} = \{V \in \mathcal{P}_k : \ \mathrm{tr}(V) = k\})$ and generating variate R if X has the same distribution as $\theta + RV^{1/2}U$, where U is uniformly distributed over $\mathcal{S}^{k-1}$ and is independent of the nonnegative random variable R. The shape V is identifiable as soon as $P[\{\theta\}] < 1$. Note that shape matrices may alternatively be normalised in such a way that they have determinant one or have an upper left entry equal to one; see, e.g., Paindaveine, 2008).

We also define a *scale functional* as a mapping $S : \mathcal{P}_k \to \mathbb{R}_0^+$ that is (i) normalised ($S(I_k) = 1$), (ii) homogeneous ($S(cA) = cS(A)$ for any $c > 0$), and (iii) monotone (if $A, B \in \mathcal{P}_k$ are such that $B - A$ positive semidefinite, then $S(B) - S(A) \geq 0$).

Shape is the parameter of interest in many multivariate statistics problems, including principal component analysis (PCA), canonical correlation analysis (CCA), testing for sphericity, etc. In PCA, for instance, since V is proportional to the covariance matrix Σ of X (when it exists), both V and Σ provide the same principal directions and the same proportions of explained variance. Moreover, under infinite second-order moments, shape remains well-defined (unlike Σ) and still fixes the elliptical geometry of the distribution that allows to conduct PCA.

The paramount importance of shape in multivariate statistics explains the huge litera-

[2]To emphasise the fact that the depth introduced in this chapter is not of a halfspace nature, we will not use the same notation $HD_P^{\mathrm{loc}}(\cdot)$ as before but rather a more generic one: $D(\cdot, P)$.

ture dedicated to M-estimation for shape/scatter parameters; see, among many others, Tyler (1987), Kent & Tyler (1988, 1991), Dudley *et al.* (2009), or the survey paper Dümbgen *et al.* (2015). It also makes it desirable to extend the concept outside the elliptical setup. Letting $U_{\theta,V}$ be the multivariate sign defined as $V^{-1/2}(X-\theta)/\|V^{-1/2}(X-\theta)\|$ if $X \neq \theta$ and as 0 otherwise (throughout, $A^{-1/2}$ is the inverse of $A^{1/2}$), Tyler (1987) defined the shape of $P = P^X$ as the matrix $V (\in \mathcal{P}_{k,\mathrm{tr}})$ satisfying

$$E\big(W_{\theta,V}\big) = 0, \qquad \text{with } \; W_{\theta,V} = \mathrm{vec}\big(U_{\theta,V}U'_{\theta,V} - \tfrac{1}{k}I_k\big), \tag{1.2}$$

where vec stacks the columns of a matrix on top of each other and where I_k denotes the k-dimensional identity matrix. If P does not charge any hyperplane containing θ, then (1.2) admits a unique solution $V (\in \mathcal{P}_{k,\mathrm{tr}})$ that agrees with the true shape if P is elliptical with location θ; see Tyler (1987), Kent & Tyler (1988) or Dümbgen (1998). In essence, (1.2) identifies the shape V making the origin equal to the expectation of $W_{\theta,V}$, that is, making the origin most central in an L_2-sense with respect to the distribution $P^{W_{\theta,V}}$ of $W_{\theta,V}$. The present chapter finds its source in the idea that, alternatively, one may define the shape of P as the matrix $V (\in \mathcal{P}_{k,\mathrm{tr}})$ making the origin most central in the, L_1, halfspace depth sense, that is, as the value of V maximising $D(0, P^{W_{\theta,V}})$. This leads to defining shape as the value of V maximising the following shape depth.

Definition 1.1 (Tyler shape depth)**.** Let $P = P^X$ be a probability measure over $\mathbb{R}^k$ and fix $V \in \mathcal{P}_{k,\mathrm{tr}}$. (i) For any $\theta \in \mathbb{R}^k$, the fixed-θ *shape depth* of V with respect to P is $D_\theta(V,P) = D(0, P^{W_{\theta,V}}) = \inf_{u \in \mathcal{S}^{k^2-1}} P\big[u'W_{\theta,V} \geq 0\big]$. (ii) The *shape depth* of V with respect to P is $D(V,P) = D_{\theta_P}(V,P)$, where θ_P is the Tukey median of P.

Whenever the argument of $D(\cdot, P)$ is a vector (resp., a shape matrix), then the notation refers to halfspace depth (resp., to Tyler shape depth). Of course, the terminology *Tyler shape depth* refers to the use of the quantity $W_{\theta,V}$ from (1.2) to define the proposed shape depth. The definition of Tyler shape depth in the unspecified location case calls for some comments. For an unspecified location, it is natural to consider the location θ and shape V satisfying both

$$E\big(U_{\theta,V}\big) = 0 \quad \text{ and } \quad E\big(W_{\theta,V}\big) = 0; \tag{1.3}$$

see Tyler (1987) and Hettmansperger & Randles (2002). In the elliptical setup, the resulting location and shape still agree with those defined above for an elliptical distribution. Here, the corresponding L_1-approach leads to considering the values of θ and V maximising $D(0, P^{U_{\theta,V}}) + \lambda D(0, P^{W_{\theta,V}})$ for some $\lambda > 0$. Interestingly, the solution does not depend on λ; the properties of halfspace depth indeed ensure that, for any V, the mapping $\theta \mapsto D(0, P^{U_{\theta,V}})$ is maximised at θ_P. Hence, for any λ, the L_1-approach identifies the location θ_P and the shape V maximising $D(0, P^{W_{\theta_P,V}})$. This justifies the "plug-in approach" adopted in Definition 1.1 for the unspecified location case. Interestingly, to the best of our knowledge, there is no formal proof that, under appropriate smoothness conditions on the underlying probability measure, there exists a solution (θ, V) of (1.3); this is why, as far as theory is concerned, the plug-in approach is adopted in the M-estimation framework; see Tyler (1987) and Hettmansperger & Randles (2002). The depth construction above has the advantage that the plug-in approach and the joint location-scatter one provide the *same* depth-based functionals.

The above concept of shape depth raises many questions: does a deepest shape matrix exist for any P? If it exists, does it coincide, under ellipticity assumptions, with the shape defined in the elliptical case? What are the properties of Tyler shape depth and of the corresponding depth regions $R_\theta(\alpha, P) = \{V \in \mathcal{P}_{k,\mathrm{tr}} : D_\theta(V, P) \geq \alpha\}$ and $R(\alpha, P) = R_{\theta_P}(\alpha, P)$? Can one perform inference based on the sample version of these concepts? Is the resulting ordering of shape matrices of interest for applications? We answer these questions in this chapter. Depth for a generic parameter has been discussed in Mizera (2002). To the best of our knowledge, however, depth for covariance or scatter matrices has only been considered in Zhang (2002), Chen *et al.* (2017) and Paindaveine & Van Bever (2017a), and only the latter work considers depth for shape matrices. x

2 Main properties

In this section, we study the main properties of the fixed-θ Tyler shape depth and of the corresponding depth regions. All proofs of results included in this chapter can be found in the second section of the Appendix. Topological statements for subsets of $\mathcal{P}_{k,\mathrm{tr}}$ and for functions defined on $\mathcal{P}_{k,\mathrm{tr}}$ will refer to the topology whose open sets are generated by balls of the form $B(V_0, r) = \{V \in \mathcal{P}_{k,\mathrm{tr}} : d(V, V_0) < r\}$, where d is the geodesic distance that is usually defined on $\mathcal{P}_k$: with the usual logarithmic mapping on $\mathcal{P}_k$, the distance d is defined through $d(V_a, V_b) = \| \log(V_a^{-1/2} V_b V_a^{-1/2})\|_F$, where $\|A\|_F = \{\mathrm{tr}(AA')\}^{1/2}$ is the Frobenius norm of A and where, for $A = \sum_{\ell=1}^k \lambda_{\ell,A} v_{\ell,A} v'_{\ell,A} \in \mathcal{P}_k$, we let $\log A = \sum_{\ell=1}^k \log(\lambda_{\ell,A}) v_{\ell,A} v'_{\ell,A}$; see, e.g., Bhatia (2007). We start with the following continuity result.

Theorem 2.1. Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Then, (i) $V \mapsto D_\theta(V, P)$ is upper semicontinuous on $\mathcal{P}_{k,\mathrm{tr}}$; (ii) the depth region $R_\theta(\alpha, P)$ is closed for any $\alpha \geq 0$; (iii) if P is absolutely continuous with respect to the Lebesgue measure, then $V \mapsto D_\theta(V, P)$ is also lower semicontinuous, hence continuous, on $\mathcal{P}_{k,\mathrm{tr}}$.

We will say that a subset R of $\mathcal{P}_{k,\mathrm{tr}}$ is bounded if and only if $R \subset B(I_k, r)$ for some $r > 0$ (the fact that the distance d actually satisfies the triangle inequality allows to restrict to balls centered at I_k). Defining $t_{\theta,P} = \sup_{u \in \mathcal{S}^{k-1}} P[u'(X - \theta) = 0]$, we say that P is smooth at θ if $t_{\theta,P} = 0$, that is, if P does not charge any hyperplane containing θ. We then have the following boundedness result.

Theorem 2.2. Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Then, the depth region $R_\theta(\alpha, P)$ is bounded and compact for any $\alpha > t_{\theta,P}$.

The main reason to adopt the geodesic distance rather than the Frobenius one $d_F(V_1, V_2) = \|V_2 - V_1\|_F$ is that, unlike $(\mathcal{P}_{k,\mathrm{tr}}, d_F)$, the metric space $(\mathcal{P}_{k,\mathrm{tr}}, d)$ is complete; see, e.g., Proposition 10 in Bhatia & Holbrook (2006). This is what allows to establish compacity in Theorem 2.2, which in turn is the main ingredient for the following result stating the existence of a deepest shape matrix.

Theorem 2.3. Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. (i) If $R_\theta(t_{\theta,P}, P)$ is non-empty, then there exists a shape $V_* \in \mathcal{P}_{k,\mathrm{tr}}$ maximising $D_\theta(V, P)$. In particular, (ii) if P is smooth at θ, then such a deepest shape V_* exists.

The deepest shape matrix V_* is a natural candidate for *the (fixed-θ) shape matrix* $V_{\theta,P}$ of P. While the previous result guarantees existence in particular for absolutely continuous probability measures, unicity is not guaranteed in general. Parallel to what is done for the Tukey median θ_P, we then define the (fixed-θ) shape matrix of P as the barycentre of the deepest shape region of P, that is, as the shape matrix $V_{\theta,P}$ satisfying

$$\operatorname{vec} V_{\theta,P} = \int_{\operatorname{vec} R_\theta(\alpha_*,P)} v\, dv \Big/ \int_{\operatorname{vec} R_\theta(\alpha_*,P)} dv, \tag{2.1}$$

where $\alpha_* = \max_V D_\theta(V,P)$. Two remarks are in order. First, the integrals in (2.1) exist and are finite since $\operatorname{vec}\mathcal{P}_{k,\mathrm{tr}}$ is a bounded subset of $\mathbb{R}^{k^2}$ ($V_{ij}^2 \leq V_{ii}V_{jj} \leq k^2$ for any $V = (V_{ij}) \in \mathcal{P}_{k,\mathrm{tr}}$). Second, the shape $V_{\theta,P}$ has maximal depth, which follows from the following convexity result.

Theorem 2.4. Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Then, (i) $V \mapsto D_\theta(V,P)$ is quasi-concave on $\mathcal{P}_{k,\mathrm{tr}}$, in the sense that

$$D_\theta((1-t)V_a + tV_b, P) \geq \min(D_\theta(V_a,P), D_\theta(V_b,P))$$

for any $V_a, V_b \in \mathcal{P}_{k,\mathrm{tr}}$ and any $t \in [0,1]$; (ii) the depth region $R_\theta(\alpha,P)$ is convex for any $\alpha \geq 0$.

This defines the (fixed-θ) shape of an arbitrary probability measure P under the extremely mild condition that $R_\theta(t_{\theta,P},P)$ is non-empty. Of course, it is important that, under ellipticity, this agrees with the elliptical definition of shape provided in the introduction. The following Fisher consistency result confirms this is the case.

Theorem 2.5. Let P be an elliptical probability measure over $\mathbb{R}^k$ with location θ_0 and shape V_0. Then, $D_{\theta_0}(V_0,P) \geq D_{\theta_0}(V,P)$ for any $V \in \mathcal{P}_{k,\mathrm{tr}}$, and, provided that $P[\{\theta_0\}] < 1$, the equality holds if and only if $V = V_0$. Letting $Y_k \sim \mathrm{Beta}(1/2,(k-1)/2)$, the maximal depth is $D_{\theta_0}(V_0,P) = (1 - P[\{\theta_0\}])P[Y_k > 1/k]$.

The only role of the constraint $P[\{\theta_0\}] < 1$ in this result is to guarantee identifiability of V_0. Note that Lemma 2 in Paindaveine & Van Bever (2017b) implies that the maximal depth in Theorem 2.5 is monotone decreasing in k as soon as $P[\{\theta_0\}]$ does not depend on k, in which case the maximal depth converges as k goes to infinity. Since Y_k has the same distribution as $Z_1^2/(\sum_{\ell=1}^k Z_\ell^2)$, where $Z = (Z_1,\ldots,Z_k)'$ is standard normal, the limit is then equal to $P[Z_1^2 > 1] \approx 0.317$. The proof of Theorem 2.5 requires the following affine-invariance/equivariance result.

Theorem 2.6. Let $P = P^X$ be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Then, for any shape matrix V, any invertible $k \times k$ matrix A and any k-vector b,

$$D_{A\theta+b}\big(V_A, P^{AX+b}\big) = D_\theta(V,P^X) \quad \text{and} \quad R_{A\theta+b}(\alpha, P^{AX+b}) = \big\{V_A : V \in R_\theta(\alpha,P)\big\},$$

where $V_A = kAVA'/\mathrm{tr}(AVA')$ is the shape matrix proportional to AVA'.

This result, which is of independent interest, shows that the fixed-θ shape depth and the corresponding regions behave well under affine transformations, hence in particular under changes of the measurement units. In the location setup, the corresponding affine-invariance/equivariance property is one of the classical requirements for depth; see Property (P1) in Zuo & Serfling (2000).

Tyler shape depth is a sign concept in the sense that it depends on the underlying random vector X only through its multivariate sign $U_{\theta,V}$. In the elliptical case, it follows that, as soon as the distribution does not charge the center of the distribution, this depth is distribution-free in the sense that it does not depend on the distribution of the underlying generating variate R, hence on the fact that the elliptic distribution is Gaussian, t, etc. More precisely, we have the following result, that plays an important role for inference based on Tyler shape depth; see Section 4.

Theorem 2.7. Let P be an elliptical probability measure over $\mathbb{R}^k$ with location θ_0 and shape V_0. Then, (i) for some $h : \mathcal{P}_{k,\mathrm{tr}} \to [0,1]$ that does not depend on V nor on P,

$$D_{\theta_0}(V,P) = (1 - P[\{\theta_0\}])\, h\left(\frac{k(V_0^{-1/2} V V_0^{-1/2})}{\mathrm{tr}(V_0^{-1}V)}\right);$$

(ii) for $k = 2$,

$$D_{\theta_0}(V,P) = (1 - P[\{\theta_0\}])\, P\left[Y_2 \geq \frac{1}{2} + \frac{1}{2}\left[1 - \det\left\{\frac{2V_0^{-1}V}{\mathrm{tr}(V_0^{-1}V)}\right\}\right]^{1/2}\right], \tag{2.2}$$

with $Y_2 \sim \mathrm{Beta}(1/2, 1/2)$.

This result shows that, while depth in the elliptical case depends on P through V_0 and $P[\{\theta_0\}]$, its dependence on $P[\{\theta_0\}]$ does not have any impact on the induced ranking of shape matrices. The explicit bivariate elliptical depth in (2.2) is compatible with all results of this section. In particular, it is easy to check that, provided $P[\{\theta_0\}] < 1$, (2.2) is uniquely maximised at $V = V_0$ and that the corresponding maximal depth is the one provided in Theorem 2.5. Boundedness of all depth regions $R(\alpha, P)$, $\alpha > 0$, can also be seen from (2.2): if $d(V, I_2)$ converges to infinity, then $d(V, V_0)$ also does, which implies that the smallest eigenvalue of $V_0^{-1}V$, hence also the depth in (2.2), converges to zero. This strengthens the general result in Theorem 2.2, that was only ensuring that the regions $R(\alpha, P)$, $\alpha > P[\{\theta_0\}]$, are bounded (note indeed that $t_{\theta_0,P} = P[\{\theta_0\}]$ for an elliptical probability measure).

3 Consistency results

Whenever k-variate observations $X_1, \ldots, X_n$ are available, the sample (fixed-θ) depth of a shape matrix V may simply be defined as $D_\theta(V, P_n)$, where P_n denotes the empirical probability measure associated with $X_1, \ldots, X_n$. In this section, we state a Glivenko-Cantelli-type result for this sample depth and investigate consistency of max-depth shape estimators. The Glivenko-Cantelli result is the following.

Theorem 3.1. Let P be a probability measure over $\mathbb{R}^k$ and let P_n denote the empirical probability measure associated with a random sample of size n from P. Then, for any $\theta \in \mathbb{R}^k$, $\sup_{V \in \mathcal{P}_{k,\mathrm{tr}}} |D_\theta(V, P_n) - D_\theta(V, P)| \to 0$ almost surely as $n \to \infty$.

We illustrate this result in the bivariate elliptical case associated with Theorem 2.7(ii). Figure 1 provides, for three different bivariate normal probability measures P, contour plots of

$$(V_{11}, V_{12}) \mapsto D_\theta(V, P), \qquad \text{with } V = \begin{pmatrix} V_{11} & V_{12} \\ V_{12} & 2 - V_{11} \end{pmatrix}$$

as well as the empirical contour plots obtained from a random sample of size $n = 800$ drawn from the corresponding distributions. Clearly, the results support the consistency in Theorem 3.1.

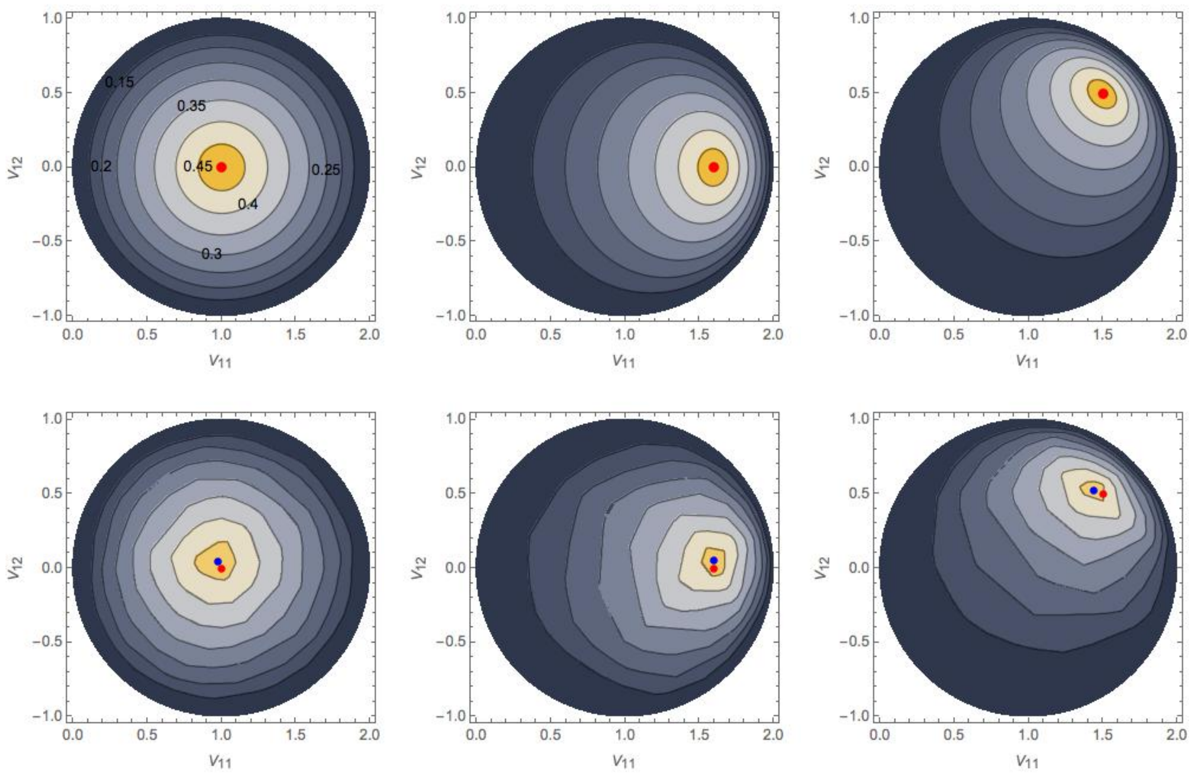

Figure 1: *(First row:) Contour plots of* $(V_{11}, V_{12}) \mapsto D_\theta(V, P)$*, for* $V = \left(\begin{smallmatrix} V_{11} & V_{12} \\ V_{12} & 2-V_{11} \end{smallmatrix}\right)$*, where* P *is bivariate normal with location* 0 *and shape* $V_A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ *(left),* $V_B \propto \left(\begin{smallmatrix} 4 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ *(center) and* $V_C \propto c\left(\begin{smallmatrix} 3 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ *(right). (Second row:) Contour plots for* $(V_{11}, V_{12}) \mapsto D_0(V, P_n)$*, where* P_n *is the empirical probability measure associated with a random sample of size* $n = 800$ *from the centered bivariate normal with shape* V_A *(left),* V_B *(center), and* V_C *(right). The "true" shapes* $V_{0,P}$ *(resp., sample deepest shapes* V_{0,P_n}*) are marked in red (resp., in blue).*

In the previous section, the shape $V_{\theta,P}$ of the probability measure P was defined as the barycentre of the collection of P-deepest shape matrices. In the empirical case, a natural estimator is of course the corresponding shape matrix V_{θ,P_n} computed from the empirical probability measure P_n associated with the sample at hand. Since empirical probability measures are not smooth at θ, existence of deepest shape matrices in the sample case does not rely on Theorem 2.3; existence, however, merely follows from the fact that the only possible values of $D_\theta(V, P_n)$ are of the form ℓ/n, $\ell = 0, 1, \ldots, n$.

Theorem 3.2. Let P be a probability measure over $\mathbb{R}^k$ and let P_n denote the empirical probability measure associated with a random sample of size n from P. Fix $\theta \in \mathbb{R}^k$ and assume that $R_\theta(t_{\theta,P}, P)$ is non-empty. Then, $d(V_{\theta,P_n}, V_{\theta,P}) \to 0$ almost surely as $n \to \infty$.

The only role of the assumption that $R_\theta(t_{\theta,P}, P)$ is non-empty is to guarantee the existence of the (fixed-θ) shape matrix $V_{\theta,P}$ of P. Again, this assumption is fulfilled in particular if P is smooth at θ. Figure 1 also supports Theorem 3.2 since, in each sample considered, the sample deepest shape is close to its population counterpart.

4 Depth-based tests for shape

We turn to hypothesis testing and focus on one-sample shape testing in the elliptical model. More specifically, based on a random sample $X_1, \ldots, X_n$ from a k-variate elliptical distribution with known location θ and unknown shape V, we want to test $\mathcal{H}_0 : V = V_0$ against $\mathcal{H}_1 : V \neq V_0$ at level $\alpha \in (0, 1)$, where $V_0 \in \mathcal{P}_{k,\mathrm{tr}}$ is fixed ($V_0 = I_k$ provides the problem of testing sphericity against elliptical alternatives). From Theorem 2.5, a natural depth-based test, ϕ_D say, rejects the null whenever $T_{\theta,n} = D_\theta(V_0, P_n) < t_{\alpha,n}$, where P_n is the empirical distribution associated with the random sample at hand and where $t_{\alpha,n}$ denotes the null α-quantile of $T_{\theta,n}$. Under the mild assumption that P does not charge the center of the distribution, $T_{\theta,n}$ is distribution-free under the null, which allows to approximate $t_{\alpha,n}$, via $\hat{t}_{\alpha,n}$ say, arbitrarily well through simulations. The resulting depth-based test then rejects the null $\mathcal{H}_0$ at level α whenever $T^{(n)} < \hat{t}_{\alpha,n}$.

Table 1 provides estimated values of the quantiles for different values of n (namely, $n = 50$, 200, 500, 10^3 and 10^4) and $\alpha = 0.01$, 0.025, 0.05, 0.1 and 0.2, both for $d = 2$ and $d = 3$. All (finite-sample) critical values were estimated on the basis of $5{,}000$ independent samples. Note that, due to the discreteness of the distribution of $Q^{(n)}_{S,D}$, some randomisation may be necessary (particularly so for small n) in order to achieve exact α-level.

Table 1: *Estimated critical values $\hat{t}_{\alpha,n}$ from $m = 5{,}000$ independent k-dimensional standard normal random samples ($k = 2, 3$), for various values of the nominal level α and sample size n.*

	$k = 2$ ($k = 3$)				
$\alpha \setminus n$	50	200	500	$1{,}000$	$10{,}000$
0.01	0.26 (0.14)	0.38 (0.28)	0.422 (0.334)	0.443 (0.360)	0.4824 (0.4031)
0.025	0.28 (0.16)	0.385 (0.29)	0.428 (0.340)	0.448 (0.364)	0.4837 (0.4044)
0.05	0.30 (0.18)	0.395 (0.295)	0.434 (0.344)	0.452 (0.368)	0.4848 (0.4056)
0.1	0.32 (0.18)	0.405 (0.305)	0.438 (0.350)	0.457 (0.371)	0.4862 (0.4070)
0.2	0.34 (0.20)	0.415 (0.315)	0.446 (0.356)	0.462 (0.376)	0.4879 (0.4084)

We performed two simulations in the bivariate case ($k = 2$). The first one considers the problem of testing sphericity about the origin ($V_0 = I_2$, $\theta = 0$) and compares the finite-sample powers of ϕ_D with those of some competitors. For each value of $\ell = 0, 1, \ldots, 6$ we generated $M = 3{,}000$ independent random samples X_i, $i = 1, \ldots, n$ of size $n = 500$ from the normal with location $\theta = 0$ and shape

$$V_{\ell,\xi} = I_2 + \ell\xi \begin{pmatrix} 1 & 0.5 \\ 0.5 & -1 \end{pmatrix}$$

and from the corresponding elliptical Cauchy. The value $\ell = 0$ corresponds to the null, whereas $\ell = 1, \ldots, 6$ provide increasingly severe alternatives. We took $\xi = 0.035$ and 0.045 for the normal and Cauchy samples in order to obtain roughly the same rejection frequencies in both cases.

For each sample, we carried out five (fixed-θ) tests at nominal level 5%: (i) the test ϕ_D rejecting the null if $T_{\theta,500} > \hat{t}_{\alpha,500} = 0.434$; (ii) the Gaussian test from John (1972)—more precisely, its extension to elliptical distributions with finite fourth-order moments from Hallin & Paindaveine (2006b); (iii)-(iv) the sign test and van der Waerden signed-rank test from the same paper; (v) the test based on the MCD$_\gamma$ shape estimator with $\gamma = 0.8$ (to achieve a good balance between efficiency and robustness) from Paindaveine & Van Bever (2014). The tests (ii)-(v) were performed based on their asymptotic null distribution. The resulting rejection frequencies in Figure 2 reveal that the depth-based test ϕ_D performs very similarly to (although it may be slightly dominated by) the sign test in (ii), which is in line with the sign nature of ϕ_D. Consequently, ϕ_D performs very well under heavy tails, where it beats all other tests. As expected, the MCD test shows low empirical powers and the Gaussian test collapses under heavy tails.

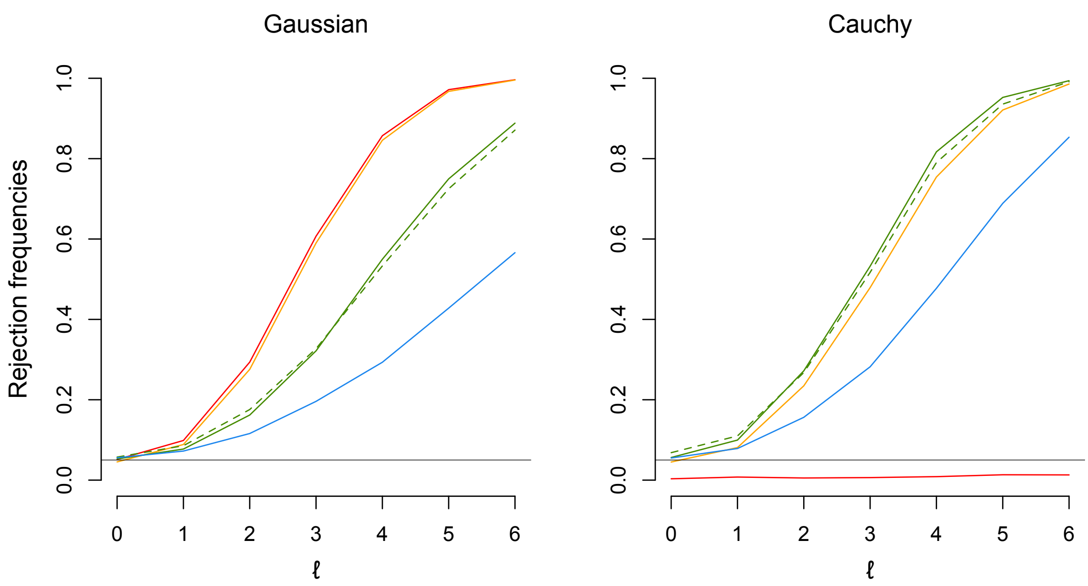

Figure 2: *Rejection frequencies, under bivariate normal (left) and elliptical Cauchy (right) densities, of five tests of sphericity: the Gaussian test (red), the van der Waerden signed-rank test (orange), the sign test (green), the depth-based test (dashed green), and the MCD-based test (blue). Results are based on 3,000 replications and the sample size is* $n = 500$. *See Section 4 for details.*

The second simulation compares the five tests above in terms of robustness when testing $\mathcal{H}_0 : V = V_0$, with $V_0 = \text{diag}(2, 1/2)$ and specified location $\theta = 0$. We focused on "level robustness" (He *et al.*, 1990) under various contaminations. We considered mixture distributions $P^{X_{(\eta)}} = (1 - \eta)P^X + \eta P^Y$, with $\eta = 0$ (no contamination), 0.025, 0.05, 0.1, 0.2, 0.25 or 0.3 (increasingly severe contaminations). Here, X is a bivariate, normal or elliptical Cauchy, null random vector. The bivariate random vector Y determines the contamination pattern and was chosen as follows: (i) *non-uniform directional contamination*: Y has the same distribution as the vector obtained by rotating X about the origin by an angle $\pi/4$ radiant; (ii) *uniform directional contamination*: Y has the same elliptical distribution as X but for the fact that its shape is $V = I_2$; (iii) *radial and uniform directional contamination*: Y is obtained by multiplying by four the vector Y in (ii). The uncontaminated distribution P^X puts more mass along the horizontal axis. In (i), the contamination is directional and typically shows along the main bisector, whereas the contamination in (ii) is uniformly distributed over the unit circle. As for (iii), the contamination combines the directional feature of (ii) with a radial outlyingness.

For each combination of a distribution type (normal or Cauchy), of a contamination pattern ((i)-(iii)), and of a contamination level η ($\eta = 0$, 0.025, 0.05, 0.1, 0.2, 0.25 or 0.3), we generated $3,000$ independent random samples $X_{(\eta)i}$, $i = 1, \ldots, n$ of size $n = 200$. The resulting rejection frequencies are plotted in Figure 3 and reveal the very good robustness of the depth-based test ϕ_D. In particular, ϕ_D always dominates its sign-based competitor. The MCD test seems to dominate ϕ_D in some configurations but, as shown in the first simulation, exhibits very low finite-sample powers. Finally, radial outliers strongly affect the Gaussian and van der Waerden tests.

5 Comparison with parametric depth

In this section, we shortly comment on how Tyler shape depth compares with the generic parametric depth concept proposed in Mizera (2002). Consider a random k-vector X with a distribution $P = P_{\vartheta_0}$ from the parametric family $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta \subset \mathbb{R}^\ell\}$ and let $\vartheta \mapsto F_\vartheta(X)$ be a measure of fit of the parameter value ϑ for an observation X. The Mizera (2002) *tangent depth* of ϑ with respect to $P = P^X$ is then $TD(\vartheta, P) = D(0, P^{\nabla_\vartheta F_\vartheta(X)})$, where $P^{\nabla_\vartheta F_\vartheta(X)}$ stands for the distribution of $\nabla_\vartheta F_\vartheta(X)$ under P. As advocated, e.g., in Mizera & Müller (2004) and Müller (2005), a likelihood-guided approach consists in taking $F_\vartheta(X) = \log L_\vartheta(X)$, where $L_\vartheta(X)$ is the likelihood of X under P_ϑ. For location and shape parameters, it is natural to consider the elliptical likelihood

$$x \mapsto L_{\theta,V,g}(x) = \frac{c_{k,g}}{(\det V)^{1/2}}\, g\big(\{(x-\theta)'V^{-1}(x-\theta)\}^{1/2}\big), \tag{5.1}$$

where $\theta \in \mathbb{R}^k$, $V \in \mathcal{P}_{k,\mathrm{tr}}$, $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is a smooth monotone decreasing function and $c_{k,g}$ is a normalising constant. Irrespective of V and g, the resulting tangent location depth $TD(\theta, P) = D(0, P^{\nabla_\theta \log L_{\theta,V,g}(X)})$ coincides with the halfspace depth of θ with respect to P.

Describing the corresponding tangent shape depth requires the following notation. Let $\mathrm{vech}(A)$ be the vector stacking the upper-diagonal entries of A on top of each other and write $\mathrm{vech}_0(A)$ for the d_k-vector ($d_k = k(k+1)/2 - 1$) obtained by depriving $\mathrm{vech}(A)$ of its first component. Further, let B_k be the $d_k \times k^2$ matrix such that $B_k'(\mathrm{vech}_0 A) = \mathrm{vec}\, A$ for any $k \times k$ symmetric matrix A with trace zero. Writing ∇_V for the gradient with respect to $\mathrm{vech}_0(V)$, the tangent shape depth of V with respect to $P = P^X$ is then

$$TD_{\theta,g}(V, P) = D(0, P^{\nabla_V \log L_{\theta,V,g}(X)}),$$

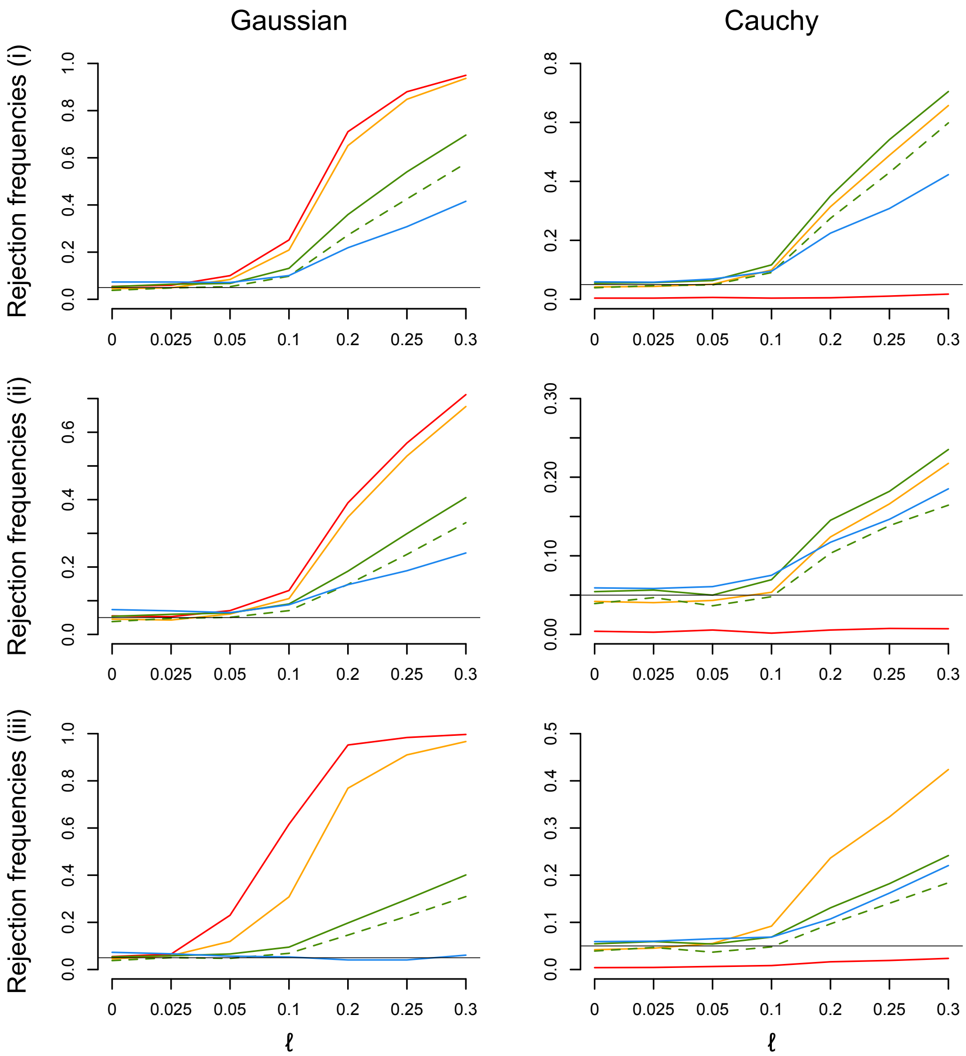

Figure 3: *Null rejection frequencies, as a function of the contamination level η, of the same five tests (using the same colours) as in Figure 2, under bivariate normal (left) and elliptical Cauchy (right) densities. The labels (i)-(iii) refer to the three contaminations patterns considered; see Section 4 for details. Results are based on 3,000 replications and the sample size is* $n = 200$.

where the score (see Hallin & Paindaveine, 2006a or Paindaveine, 2008)

$$\nabla_V \log L_{\theta,V,g}(X) = \tfrac{1}{2}\, B_k(V^{\otimes 2})^{-1/2}\, \mathrm{vec}\big(\varphi_g\big(d_{\theta,V}\big) d_{\theta,V} U_{\theta,V} U'_{\theta,V} - \tfrac{1}{k} I_k\big)$$

involves the Mahalanobis distance $d_{\theta,V} = \{(X-\theta)'V^{-1}(X-\theta)\}^{1/2}$. Tangent shape depth is much less satisfactory than its location counterpart : first, it depends on g in (5.1); for instance, a Gaussian likelihood and a t_ν likelihood will provide different tangent shape depths. Second, more importantly, the tangent deepest shape is not Fisher-consistent under ellipticity, irrespective of g, as the following example shows. Let X be a bivariate normal random vector with location $\theta_0 = 0$ and shape $V_0 = \mathrm{diag}(3/4, 5/4)$. Then it can be showed (see Lemma B.7 in the Appendix) that, even when considering the tangent shape depth associated with the "true" location θ and function g (i.e., the one obtained with $\theta = \theta_0$ and $g(r) = g_0(r) = \exp(-r^2/2)$ in (5.1)), one has $TD_{\theta_0,g_0}(V_0, P^X) < TD_{\theta_0,g_0}(I_2, P^X)$, which implies that the true shape V_0 does not maximise $V \mapsto TD_{\theta_0,g_0}(V, P^X)$. This clearly disqualifies tangent depth for shape parameters. For the sake of illustration, Figure 4 plots (an estimated version of) $TD_{\theta_0,g_0}(V_a, P^X)$, for $V_a = \mathrm{diag(a, 2-a)}$, as a function of $a \in (0,2)$. Estimation was obtained as follows: we generated $M = 100$ mutually independent random samples from the distribution P_X described above. Then, for every value $a_i = i/100$, with $i = 1,\ldots,199$, we estimated $TD_{\theta_0,g_0}(V_a, P^X)$ by averaging over the $M = 100$ samples available the respective sample depths (obtained by taking the sample halfspace depth of the score vectors). Figure 4 indeed confirms that Fisher consistency does not hold in that instance.

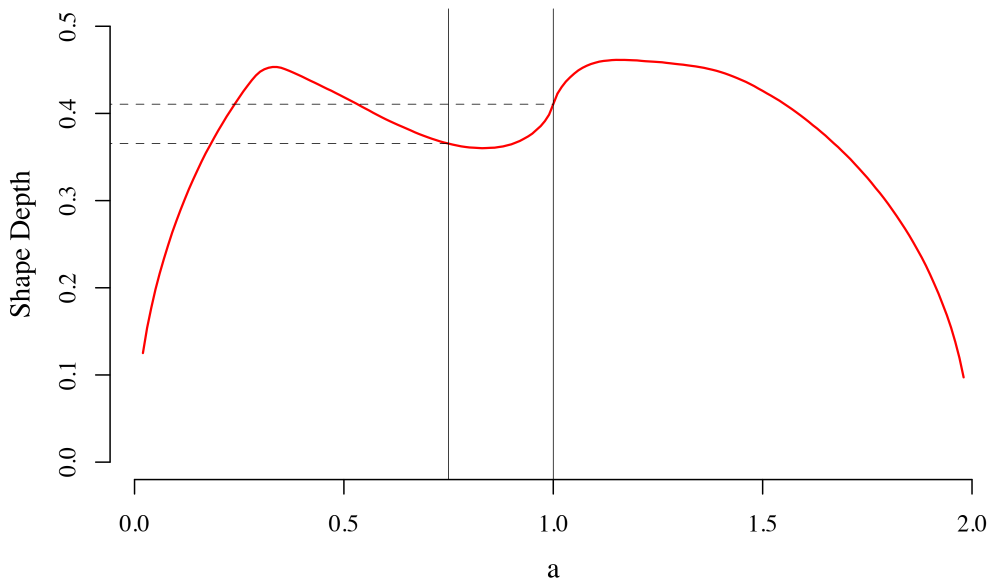

Figure 4: *Plot of (estimated) $TD_{\theta_0,g_0}(V_a, P^X)$ as function of $a \in (0,2)$, where X is bivariate gaussian with true shape V_{a_0}, for $a_0 = 3/4$.*

6 The unspecified location case and a real-data example

The previous sections focused on the fixed-θ shape depth $D_\theta(V,P)$. Most of the result extend, with only minor modifications (if any), to the unspecified-location shape depth $D(V,P) = D_{\theta_P}(V,P)$ from Definition 1.1. Theorems 2.1 to 2.4 hold for any fixed θ and their unspecified-θ versions are simply obtained by substituting θ_P for θ throughout. In particular, the existence of an unspecified-location deepest shape matrix is guaranteed if P is smooth at θ_P, or, more generally, if $R(t_{\theta_P,P}, P)$ is non-empty. The same construction allows to identify a unique representative V_P of the collection of deepest shape matrices. Theorem 2.5 and Theorem 2.7 also readily extend to the unspecified-location case since $\theta_P = \theta_0$ for any elliptical probability measure P with location θ_0. In particular, if P is elliptical with shape V_0, then the unspecified-θ shape depth $D(V,P)$ is uniquely maximised at $V = V_0$ (as soon as the distribution is not degenerate at a single point). In view of the affine equivariance of θ_P (i.e., $\theta_{P^{AX+B}} = A\theta_{P^X} + b$), the affine-invariance/equivariance properties

$$D\big(V_A, P^{AX+b}\big) = D(V, P^X) \quad \text{and} \quad R(\alpha, P^{AX+b}) = \big\{V_A : V \in R(\alpha, P)\big\}$$

directly follow from Theorem 2.6 (see this result for the definition of V_A). As a matter of fact, only the consistency results in Theorems 3.1-3.2 require stronger assumptions (absolute continuity) in the unspecified-location case compared to the specified-location one. More precisely, we have

Theorem 6.1. Let P be a probability measure over $\mathbb{R}^k$ that is absolutely continuous with respect to the Lebesgue measure and let P_n denote the empirical probability measure associated with a random sample of size n from P. Then, (i) $\sup_{V \in \mathcal{P}_{k,\mathrm{tr}}} |D(V, P_n) - D(V, P)| \to 0$ almost surely as $n \to \infty$; (ii) $d(V_{P_n}, V_P) \to 0$ almost surely as $n \to \infty$.

We illustrate the use of the unspecified-location Tyler shape depth on a real data example. For each trading day between February 1st, 2015 and February 1st, 2017, we collected every five minutes the Nasdaq Composite and S&P500 stock indices and computed their returns, that is, the differences between two consecutive index values. The returns on a given day form a bivariate dataset of usually 78 observations. The exact number of observations per day varies due to some missing values and days with less than 70 bivariate returns were discarded. The resulting dataset comprises $n = 38489$ observations distributed over $D = 478$ trading days.

The analysis conducted here studies the joint behavior of the bivariate returns. Its goal is to determine which trading days present an atypical pattern, different from the "global" behavior of the volatility. Of course, an important source of atypicality is associated with the overall scale of the bivariate returns that alternate between periods of high and low volatility. Such deviations, however, can easily be detected by comparing, e.g., the trace of any scatter measure on intraday data with that on the whole dataset. Therefore, we rather focus on detecting atypicality in the shape of the joint volatility. In other words, we aim at detecting days for which the ratios of the marginal volatilities or the correlation between the returns much deviate from their global behavior.

To this end, let $\hat{V}_{\mathrm{full}}$ denote the unspecified-location Tyler shape M-estimate computed from the full collection of n returns, that is, (the shape part in) the solution of the empirical version of (1.3); see Tyler (1987) or Hettmansperger & Randles (2002). For each day $d = 1, \ldots, D$, we evaluated the depth $D(\hat{V}_{\mathrm{full}}, P_d)$ of the global shape estimate with respect to the empirical distribution P_d of the bivariate returns on day d. The reason why we base this measure of (a)typicality on $\hat{V}_{\mathrm{full}}$ is twofold. First, $\hat{V}_{\mathrm{full}}$ is a robust shape estimate which takes into account outliers that occurs naturally in the dataset (returns at the beginning of each trading period are notoriously more volatile and should be downweighted in the shape estimation procedure). Second, $\hat{V}_{\mathrm{full}}$ is very deep in the global series of returns (denoting as P_{full} the empirical distribution of the full collection of bivariate returns, we have $D(\hat{V}_{\mathrm{full}}, P_{\mathrm{full}}) = 0.4965$), hence is an excellent proxy for the (global) deepest shape matrix, whose computation seems to be a difficult task.

The left panel of Figure 5 presents the depth values $D(\hat{V}_{\mathrm{full}}, P_d)$ as a function of $d = 1, \ldots, D$. Vertical lines mark major events affecting the shape of the volatility, while the two greyed rectangles cover two periods during which the markets notoriously knew some atypical returns. The first period follows the devaluation of the Yuan on August 11th, 2015 which saw rapid changes in the stock markets, including large devaluations on the "Black Monday" of

August 24th (marked in orange). The second period covers the beginning of 2016, which was hit by slump in oil prices, making stocks relying on oil very volatile compared to non oil-based ones. This resulted in atypical shape behavior during the fortnight spanning January 22 - February 9 (this last day is marked in blue, and is known to have the sharpest loss for the S&P500 index). The other events are (3) the decision of the European Central Bank on March 10th, 2016 (in green) to extend quantitative easing thereby slashing interest rates (known to have had a significant positive impact on both Nasdaq and S&P500, but more pronounced for the latter), (4) the positive impact on the financial stocks following Fed officials' comments on the possibility of rate hike made on May 27, 2016 (in red), (5) the slump in the S&P500 Futures prices on August 15th, 2016 (in purple), and (6) the aftermath of Donald Trump's election at the US presidency on November 9th (in teal). All events are associated with days known to have atypical volatilities and are seen to have a low shape depth value. Some other major financial events – such as OPEC refusing to reduce oil production in early 2015 or the aftermath of the Brexit vote on June 24, 2016 (midway between events (4) and (5)) – had an effect on the overall size of the bivariate returns but not on their shape, which explains that the corresponding days are not flagged as overly atypical by Tyler shape depth.

For the sake of comparison, we also computed the halfspace shape depth $HD(\hat{V}_{\text{full}}, P_d)$ (see Section 8 in Chapter I) of the global estimate for each day d. The right panel of Figure 5 provides the plot of $D(\hat{V}_{\text{full}}, P_d)$ versus $HD(\hat{V}_{\text{full}}, P_d)$ for $d = 1, \ldots, D$. The plot shows a clear positive association (correlation between the two variables is 0.6413). The following two remarks are, however, in order. First, halfspace shape depth values seem to have a higher concentration than Tyler's. This is due to the fact that the former maximises a concept of scatter depth in scale and has, possibly, more leeway to find scatter estimates better suited to the data. Indeed, a decrease in volatility in one of the marginals might be balanced by considering a scatter with a smaller scale and hence keep a large depth value. A byproduct of this is the fact that, when evaluating halfspace shape depth, the (difficult) maximisation step in scale seems to be crucial in correctly computing the depth ranking of the data (small deviations can indeed cause changes in this ranking). More importantly, while events (1) to (3) receive low depth with respect to both concepts, only Tyler shape depth succeeds in flagging days associated with events (4) to (6) as "outlying".

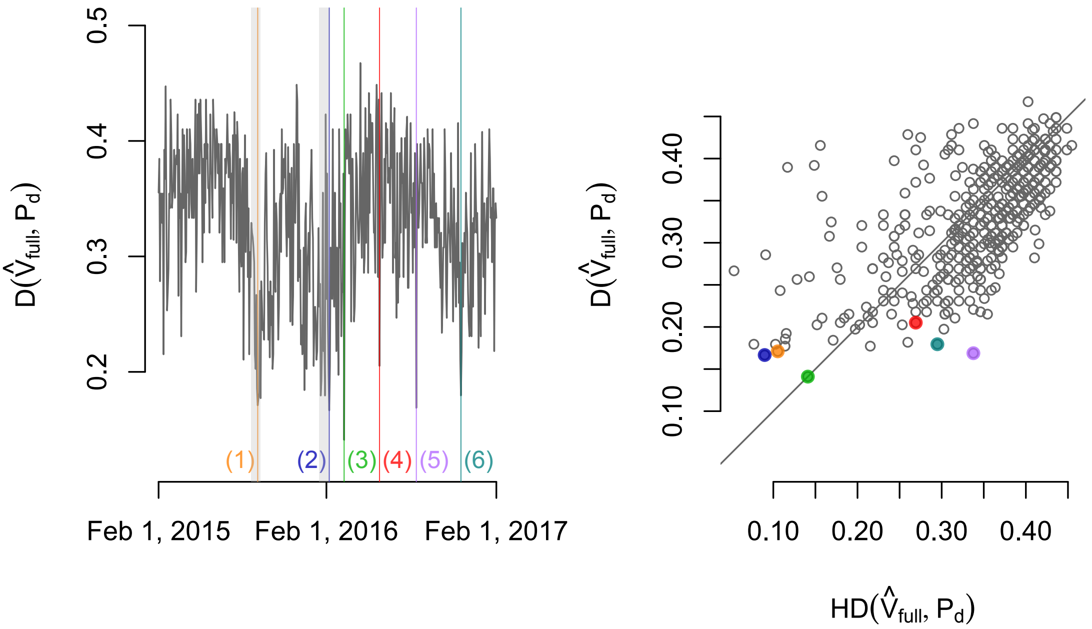

Figure 5: *(Left:) Plot of $D(\hat{V}_{\text{full}}, P_d)$ as a function of d. Events (1) to (6) are described in Section 6. (Right:) Plot of $D(\hat{V}_{\text{full}}, P_d)$ vs $HD(\hat{V}_{\text{full}}, P_d)$ for each trading day d. Events from the left panel are highlighted using the same colour.*

Final comments and perspectives

In this manuscript, we developed and studied several depth notions for dispersion parameters. The first one, $HD_P(\Sigma)$ (and the related shape and concentration concepts) is based on halfspace considerations. The second, $D_\theta(V, P)$ is restricted to shapes and parallels Tyler's M estimator of shape while allowing comparison between various shapes. The geometric structure of the parameter spaces played a crucial role in understanding the properties of the depth regions and, in turn, of the depth functions. In particular, Riemannian topology was a key ingredient in proving existence of deepest scatters, shape and concentration matrices. We close this thesis with several comments on the various concepts introduced.

In Chapter I, we thoroughly investigated the structural properties of a concept of scatter halfspace depth linked to those proposed in Zhang (2002) and Chen *et al.* (2017). While we tried doing so under minimal assumptions, alternative scatter halfspace depth concepts may actually require even weaker assumptions, but they typically would make the computational burden heavier in the sample case. As an example, one might alternatively define the scatter halfspace depth of $\Sigma(\in \mathcal{P}_k)$ with respect to P as

$$HD_P^{\mathrm{sc,alt}}(\Sigma) = \sup_{\theta \in \mathbb{R}^k} HD_{P,\theta}^{\mathrm{sc}}(\Sigma), \tag{6.1}$$

where $HD_{P,\theta}^{\mathrm{sc}}(\Sigma)$ is the scatter halfspace depth associated with the constant location functional at θ. This alternative scatter depth concept satisfies a uniform consistency result such as the one in Theorem 2.2 without any condition on P, whereas the scatter halfspace depth $HD_{P,T}^{\mathrm{sc}}(\Sigma)$ in (2.1) requires that P is smooth (see Theorem 2.2). In the sample case, however, evaluation of $HD_{P_n}^{\mathrm{sc,alt}}(\Sigma)$ is computationally much more involved than $HD_{P_n,T}^{\mathrm{sc}}(\Sigma)$. Alternative concentration and shape halfspace depth concepts may be defined along the same lines and will show the same advantages/disadvantages. compared to those proposed in this chapter.

Another possible concept of scatter halfspace depth bypasses the need to choose a location functional T by exploiting a pairwise difference approach; see Zhang (2002) and Chen *et al.* (2017). In our notation, the resulting scatter depth of Σ with respect to $P = P^X$ is

$$HD_P^{\mathrm{sc,U}}(\Sigma) = HD_{P^{X-\tilde{X}},0}^{\mathrm{sc}}(\Sigma), \tag{6.2}$$

where $\tilde{X}$ is an independent copy of X and where 0 denotes the origin of $\mathbb{R}^k$. On one hand, the sample version of (6.2) is a U-statistic of order two, which will increase the computational burden compared to the sample version of (2.1). On the other hand, uniform consistency results for (6.2) (which here follow from Glivenko-Cantelli results for U-processes, such as the one in Corollary 3.3 from Arcones & Giné, 1993) will again hold without any assumption on P, which is due to the fact that, as already mentioned, the smoothness assumption in Theorem 2.2 is superfluous when a constant location functional T is used. At first sight, thus, the pros and cons for (6.2) are parallel to those for (6.1), that is, weaker distributional assumptions are obtained at the expense of computational ease. However, (6.2) suffers from a major disadvantage: it does not provide Fisher consistency at the elliptical model (see (Q2) in Section 5). This results from the fact that if $P = P^X$ is elliptical with location θ and scatter Σ, then $P^{X-\tilde{X}}$ is elliptical with location 0 and scatter $c_P\Sigma$, where the scalar factor c_P depends on the type of elliptical distribution: for multinormal and Cauchy elliptical distributions, e.g., $c_P = 2$ and 4, respectively, so that if one replaces $X - \tilde{X}$ with $(X - \tilde{X})/\sqrt{2}$

to achieve Fisher consistency at the multinormal, then Fisher consistency will still not hold at the Cauchy. Actually, the maximiser of $HD_P^{\mathrm{sc,U}}(\Sigma)$ is useless as a measure of scatter for the original probability measure P, as its interpretation requires knowing which type of elliptical distribution P is. This disqualifies the pairwise difference scatter depth, as well as the companion concentration depth. Note, however, that the corresponding shape depth will not suffer from this Fisher consistency problem since the normalisation of scatter matrices into shape matrices will get rid of the scalar factor c_P.

As both previous paragraphs suggest and as it is often the case with statistical depth, computational aspects are key for the application of the proposed depths. Evaluating (good approximations of) the scatter halfspace depth $HD^{\mathrm{sc}}_{P_n,T}(\Sigma)$ of a given Σ can of course be done for very small dimensions $k = 2$ or 3 by simply sampling the unit sphere $\mathcal{S}^{k-1}$. Even for such small dimensions, however, computing the halfspace deepest scatter is non-trivial: while scatter halfspace depth relies on a low-dimensional (that is, k-dimensional) projection-pursuit approach, identifying the halfspace deepest scatter indeed requires exploring the collection of scatter matrices $\mathcal{P}_k$, that is of higher dimension, namely of dimension $k(k+1)/2$. Fortunately, the fixed-location scatter halfspace depth — hence, also its T-version proposed here after appropriate centering of the observations — can be computed in higher dimensions through the algorithm proposed in Chen *et al.* (2017), where the authors performed simulations requiring to compute the deepest scatter matrix for dimensions and sample sizes as large as 10 and 2000, respectively. Their implementation of this algorithm is available as an R package at `https://github.com/ChenMengjie/DepthDescent`.

The concept of scatter halfspace depth also makes sense when the parameter space is the compactification of $\mathcal{P}_k$, that is, is the collection $\overline{\mathcal{P}}_k$ of $k \times k$ symmetric positive *semi*-definite matrices. Interestingly, it is actually easier to investigate the properties of scatter halfspace depth over $\overline{\mathcal{P}}_k$ than over $\mathcal{P}_k$. The F-continuity and F-boundedness results in Theorems 3.1-3.2 extend, mutatis mutandis, to $\overline{\mathcal{P}}_k$. Unlike $(\mathcal{P}_k, d_F)$, the metric space $(\overline{\mathcal{P}}_k, d_F)$ is complete, so that the regions $R^{\mathrm{sc}}_{P,T}(\alpha)$ are then F-compact for any $\alpha > 0$. Consequently, a trivial adaptation of the proof of Theorem 4.3 allows to show that there *always* exists a halfspace deepest scatter matrix in $\overline{\mathcal{P}}_k$. It is fortunate that these neat results can be established by considering the F-distance only, as the geodesic distance, that is unbounded on $\overline{\mathcal{P}}_k \times \overline{\mathcal{P}}_k$, could not have been considered here. Of course, in many applications, $\mathcal{P}_k$ remains the natural parameter space since many multivariate statistics procedures will require inverting scatter matrices. In such applications, it will be of little help to practitioners that the deepest halfspace scatter matrix belongs to $\overline{\mathcal{P}}_k \setminus \mathcal{P}_k$, which explains why our detailed investigation focusing on $\mathcal{P}_k$ is of key importance.

Perspectives for future research are rich and diverse. The proposed halfspace depth concepts for scatter, concentration and shape can be extended to other scatter functionals of interest. In particular, halfspace depths that are relevant for PCA could result from the "profile depth" approach in Section 7. For instance, the T-*"first principal direction" halfspace depth* of $\beta(\in \mathcal{S}^{k-1})$ with respect to the probability measure P over $\mathbb{R}^k$ can be defined as

$$HD^{1^{\mathrm{st}}\mathrm{pd}}_{P,T}(\beta) = \sup_{\Sigma \in \mathcal{P}_{k,1,\beta}} HD^{\mathrm{sc}}_{P,T}(\Sigma), \quad \text{with } \mathcal{P}_{k,1,\beta} := \{\Sigma \in \mathcal{P}_k : \Sigma\beta = \lambda_1(\Sigma)\beta\}.$$

The halfspace deepest first principal direction is a promising robust estimator of the true underlying first principal direction, at least under ellipticity. Obviously, the depth of any other principal direction, or the depth of any eigenvalue, can be defined accordingly. Another direction of research is to explore inferential applications of the proposed depths. Clearly, point estimation is to be based on halfspace deepest scatter, concentration or shape matrices; Chen *et al.* (2017) partly studied this already for scatter in high dimensions. Hypothesis testing is also of primary interest. In particular, a natural test for $\mathcal{H}_0 : \Sigma = \Sigma_0$, where $\Sigma_0 \in \mathcal{P}_k$ is fixed, would reject the null for small values of $HD^{\rm sc}_{P_n,T}(\Sigma_0)$. For shape matrices, a test of sphericity would similarly reject the null for small values of $HD^{{\rm sh},S}_{P_n,T}(I_k)$. These topics are beyond the scope of this manuscript.

We conclude with a comparison between the shape depth concepts introduced in both chapters. In its specified-θ version, halfspace shape depth is obtained as $HD_\theta(V,P) = \sup_{\sigma^2>0} HD^{\rm sc}_\theta(\sigma^2 V, P)$, where $HD^{\rm sc}_\theta(\cdot, P)$ is the companion concept of halfspace depth for (unnormalised) scatter matrices. The halfspace shape depth $HD_\theta(V,P)$ therefore requires a (delicate) maximisation in σ^2 of the halfspace scatter depth, that itself requires a projection-pursuit optimisation in $\mathbb{R}^k$. In comparison, Tyler shape depth has the advantage to be intrinsically a depth for shape matrices. A possible drawback of Tyler shape depth, however, is that it in principle requires to evaluate halfspace (location) depth in $\mathbb{R}^{k^2}$, which is computationally prohibitive even for relatively small dimensions k. Interestingly, it can be showed that Tyler shape depth only requires evaluating halfspace depth in $\mathbb{R}^{d_k}$ (recall that we let $d_k = k(k+1)/2 - 1$ above). This is proved in Theorem B.1, which closes the Appendix.

From a computational point of view, thus, Tyler shape depth competes well for $k = 2$ and 3 with its halfspace counterpart; the latter, however, would dominate Tyler shape depth in this respect for larger dimensions k. Most importantly, a strong advantage of Tyler shape depth over its halfspace competitor is its distribution-freeness (Theorem 2.7), which results from the sign nature of the concept. As illustrated in Section 4 of Chapter II, distribution-freeness allows to perform inference on Tyler shape depth in the elliptical framework. This cannot be done with the halfspace shape depth that turns out to crucially depend on the type of elliptical distribution at hand.

Appendix

A Proofs from Chapter I

This first section of the Appendix collects all proofs from Chapter I. In the proofs below, we will often use the fact that

$$\begin{aligned}HD^{\rm sc}_{P,T}(\Sigma) &= \inf_u \min\big(P\big[|u'(X-T_P)| \le \sqrt{u'\Sigma u}\,\big], P\big[|u'(X-T_P)| \ge \sqrt{u'\Sigma u}\,\big]\big)\\ &= \min\Big(\inf_u P\big[|u'(X-T_P)| \le \sqrt{u'\Sigma u}\,\big], \inf_u P\big[|u'(X-T_P)| \ge \sqrt{u'\Sigma u}\,\big]\Big),\end{aligned}$$

where all infima are over the unit sphere $\mathcal{S}^{k-1}$ of $\mathbb{R}^k$.

A.1 Proofs from Section 2

Proof of Theorem 2.1. Fix $A \in GL_k$ and $b \in \mathbb{R}^k$. By using the affine-equivariance of T (that is, $T_{P_{A,b}} = AT_P + b$) and letting $u_A := A'u/\|A'u\|$, we obtain

$$\begin{aligned}&HD^{\rm sc}_{P_{A,b},T}(A\Sigma A')\\ &= \inf_{u\in\mathcal{S}^{k-1}} \min\big(P\big[|u'A(X-T_P)| \le \sqrt{u'A\Sigma A'u}\,\big], P\big[|u'A(X-T_P)| \ge \sqrt{u'A\Sigma A'u}\,\big]\big)\\ &= \inf_{u\in\mathcal{S}^{k-1}} \min\big(P\big[|u_A'(X-T_P)| \le \sqrt{u_A'\Sigma u_A}\,\big], P\big[|u_A'(X-T_P)| \ge \sqrt{u_A'\Sigma u_A}\,\big]\big)\\ &= HD^{\rm sc}_{P,T}(\Sigma),\end{aligned}$$

where the last equality follows from the fact that the mapping $u \mapsto u_A$ is a one-to-one transformation of $\mathcal{S}^{k-1}$. □

We now establish the explicit scatter halfspace depth expression (2.7) in the independent Cauchy case. For that purpose, consider again a random vector $X = (X_1, \ldots, X_k)'$ with independent Cauchy marginals. Using the fact that, for any non-negative real numbers a_ℓ, $\ell = 1, \ldots, k$, the random variables $\sum_{\ell=1}^k a_\ell X_\ell$ and $(\sum_{\ell=1}^k a_\ell) X_1$ then share the same distribution, we obtain that, denoting by $\|x\|_1 = \sum_{\ell=1}^k |x_\ell|$ the L_1-norm of $x = (x_1, \ldots, x_k)'$,

$$P\big[|u'X| \le \sqrt{u'\Sigma u}\,\big] = P\big[\|u\|_1\,|X_1| \le \sqrt{u'\Sigma u}\,\big] = 2\Psi\bigg(\frac{\sqrt{u'\Sigma u}}{\|u\|_1}\bigg) - 1,$$

where Ψ is the Cauchy cumulative distribution function. Therefore, if T is centro-equivariant, we obtain

$$\begin{aligned}HD^{\rm sc}_{P,T}(\Sigma) &= \min\bigg(2\Psi\bigg(\inf_{u\in\mathcal{S}^{k-1}} \frac{\sqrt{u'\Sigma u}}{\|u\|_1}\bigg) - 1, 2 - 2\Psi\bigg(\sup_{u\in\mathcal{S}^{k-1}} \frac{\sqrt{u'\Sigma u}}{\|u\|_1}\bigg)\bigg)\\ &= 2\min\Big(\Psi\big(\min_v \sqrt{v'\Sigma v}\,\big) - \tfrac{1}{2}, 1 - \Psi\big(\max_v \sqrt{v'\Sigma v}\,\big)\Big),\end{aligned}$$

where the minimum and maximum in v are over the unit L_1-sphere $\{v \in \mathbb{R}^k : \|v\|_1 = 1\}$. The formula (2.7) then follows from the following result.

Lemma A.1. For any $\Sigma \in \mathcal{P}_k$, the maximal and minimal values of $v'\Sigma v$ when v runs over the L_1-sphere $\{v \in \mathbb{R}^k : \|v\|_1 = \sum_{i=1}^k |v_i| = 1\}$ are $\max(\mathrm{diag}(\Sigma))$ and $1/\max_s(s'\Sigma^{-1}s)$, respectively, where $\max_s$ is the maximum over $s = (s_1, \ldots, s_k) \in \{-1, 1\}^k$.

Proof of Lemma A.1. We start with the following considerations. In dimension k, the L_1-sphere can be parametrised as

$$v_{s,t} = (s_1t_1, s_2t_2, \ldots, s_{k-1}t_{k-1}, s_k(1 - t_1 - \ldots - t_{k-1}))',$$

where $s = (s_1, \ldots, s_k) \in \{-1, 1\}^k$ and $(t_1, \ldots, t_{k-1}) \in \mathrm{Simpl}_{k-1} := \{(t_1, \ldots, t_{k-1}) : t_1 \geq 0, \ldots, t_{k-1} \geq 0, t_1 + \ldots + t_{k-1} \leq 1\}$. Clearly, symmetry of the L_1-sphere and of the function to be maximised/minimised allows to restrict to $s_k = 1$. Now, for any given s of the form $(s_1, \ldots, s_{k-1}, 1)$, consider the function

$$\begin{array}{rccl} f_s : & \mathrm{Simpl}_{k-1} & \to & \mathbb{R} \\ & t = (t_1, \ldots, t_{k-1}) & \mapsto & v'_{s,t}\Sigma v_{s,t}. \end{array}$$

This function is twice differentiable, with a gradient $\nabla f_s(t)$ whose ith component is

$$\begin{aligned} \frac{\partial}{\partial t_i} f_s(t) &= (0, ..., 0, s_i, 0, \ldots, -1)\Sigma v_{s,t} + v'_{s,t}\Sigma(0, ..., 0, s_i, 0, \ldots, -1)' \\ &= 2(0, ..., 0, s_i, 0, \ldots, -1)\Sigma v_{s,t} \end{aligned}$$

and a Hessian matrix $H_s(t)$ whose (i, j)-entry is

$$\frac{\partial^2}{\partial t_i t_j} f_s(t) = 2(0, ..., 0, s_i, 0, \ldots, -1)\Sigma(0, ..., 0, s_j, 0, \ldots, -1)'.$$

Now, for any $z \in \mathbb{R}^{k-1}$,

$$z'H_s(t)z = 2(s_1z_1, ..., s_kz_k, -z_1 - \ldots - z_k)\Sigma(s_1z_1, ..., s_kz_k, -z_1 - \ldots - z_k)' \geq 0,$$

with equality if and only if $z = 0$. Therefore, f_s is strictly convex over Simpl_{k-1}.

Let us start with the maximum. For a given s, strict convexity of f_s implies that the maximum of f_s can only be achieved at $e_{i,k-1}$, $i = 1, \ldots, k-1$, where $e_{i,\ell}$ stands for the ith vector of the canonical basis of $\mathbb{R}^\ell$, or at $0 (\in \mathbb{R}^{k-1})$. Since $f_s(e_{i,k-1}) = \Sigma_{ii}$, $i = 1, \ldots, k-1$, and $f_s(0) = \Sigma_{kk}$, it follows that the maximal value of f_s over Simpl_{k-1} is $\max(\mathrm{diag}(\Sigma))$. Since this is the case for any s, we conclude that the maximum of $v'\Sigma v$ over the L_1-sphere is itself $\max(\mathrm{diag}(\Sigma))$.

Let us then turn to the minimum. The minimum of f_s, when extended into a (still strictly convex) function defined on $\mathbb{R}^{k-1}$, is the solution of the gradient conditions

$$(0, ..., 0, s_i, 0, \ldots, -1)\Sigma v_{s,t} = 0, \qquad i = 1, \ldots, k-1.$$

Writing simply $e_k = e_{k,k}$ for the kth vector of the canonical basis of $\mathbb{R}^k$ and letting

$$S_s = \begin{pmatrix} s_1 & & & -1 \\ & \ddots & & \vdots \\ & & s_{k-1} & -1 \end{pmatrix},$$

these gradient conditions rewrite $S_s\Sigma(e_k + S_s' t) = 0$ (recall that we restricted to $s_k = 1$), and their unique solution in $\mathbb{R}^{k-1}$ is $t_s^{\min} := -(S_s\Sigma S_s')^{-1}S_s\Sigma e_k$.

It will be useful below to have a more explicit expression of $t_s^{\min}$. Note that the gradient conditions above state that $\Sigma(e_k + S_s' t_s^{\min})$ is in the null space of S_s. Since this null space is easily checked to be $\{\lambda s : \lambda \in \mathbb{R}\}$, this implies that $e_k + S_s' t_s^{\min} = \lambda\Sigma^{-1}s$ for some $\lambda \in \mathbb{R}$. Premultiplying both sides of this equation by s', we obtain $1 = \lambda s'\Sigma^{-1}s$, which yields $\lambda = 1/(s'\Sigma^{-1}s)$. Thus,

$$e_k + S_s' t_s^{\min} = \frac{1}{s'\Sigma^{-1}s}\,\Sigma^{-1}s. \tag{A.1}$$

In the first $k-1$ components, this yields (after multiplication by s_i)

$$(t_s^{\min})_i = \frac{s_i e_i'\Sigma^{-1}s}{s'\Sigma^{-1}s}, \qquad i = 1, \ldots, k-1, \tag{A.2}$$

while the kth component provides (still with $s_k = 1$)

$$1 - \sum_{i=1}^{k-1}(t_s^{\min})_i = \frac{e_k'\Sigma^{-1}s}{s'\Sigma^{-1}s} = \frac{s_k e_k'\Sigma^{-1}s}{s'\Sigma^{-1}s}\cdot \tag{A.3}$$

Strict convexity of f_s implies that its minimal value over $\mathbb{R}^{k-1}$ is $f_s(t_s^{\min})$. By using (A.1), this minimal value takes the form

$$f_s(t_s^{\min}) = v'_{s,t_s^{\min}}\Sigma v_{s,t_s^{\min}} = (e_k + S_s' t_s^{\min})'\Sigma(e_k + S_s' t_s^{\min}) = \frac{1}{s'\Sigma^{-1}s}\cdot$$

Now, consider an arbitrary sign k-vector s_* that maximises $s'\Sigma^{-1}s$ among the 2^{k-1} corresponding sign vectors s to be considered (the last component of s is still fixed to one). Assume for a moment that $t_{s_*}^{\min}$ is in the interior of Simpl_{k-1}. Since it is the minimal value of f_{s*} over $\mathbb{R}^{k-1}$, $f_{s_*}(t_{s_*}^{\min}) = 1/(s_*'\Sigma^{-1}s_*)$ of courses minimises f_{s*} over Simpl_{k-1}. Pick then another sign vector s. By construction, $1/(s_*'\Sigma^{-1}s_*)$ is smaller than or equal to $f_s(t_s^{\min}) = 1/(s'\Sigma^{-1}s)$, which, as the minimal value of f_s when extended to $\mathbb{R}^{k-1}$, can only be smaller than or equal to the minimal value of f_s over Simpl_{k-1}. Therefore, $1/(s_*'\Sigma^{-1}s_*)$ is then the minimal value of $v'\Sigma^{-1}v$ over the unit L_1-sphere.

It thus remains to show that $t_{s_*}^{\min}$ indeed belongs to the interior of Simpl_{k-1}. Equivalently (in view of (A.2)-(A.3)), it remains to show that $s_{*i}e_i'\Sigma^{-1}s_* > 0$ for $i = 1, \ldots, k$. Assume then that $s_{*\ell}e_\ell'\Sigma^{-1}s_* \leq 0$ for some ℓ. Defining s_{**} as the vector obtained from s_* by only changing the sign of its ℓth component, that is, putting $s_{**} := s_* - 2s_{*\ell}e_\ell$, we have

$$\begin{aligned} s_{**}'\Sigma^{-1}s_{**} - s_*'\Sigma^{-1}s_* &= (s_* - 2s_{*\ell}e_\ell)'\Sigma^{-1}(s_* - 2s_{*\ell}e_\ell) - s_*'\Sigma^{-1}s_* \\ &= -4s_{*\ell}e_\ell'\Sigma^{-1}(s_* - s_{*\ell}e_\ell) = -4s_{*\ell}e_\ell'\Sigma^{-1}s_* + 4e_\ell'\Sigma^{-1}e_\ell > 0, \end{aligned}$$

which contradicts the maximality property of s_*. $\square$

The proof of Theorem 2.2 requires Lemmas A.2-A.4 below. Before proceeding, we introduce the inner and outer "slabs"

$$H_{u,c}^{\text{in}} := \{x \in \mathbb{R}^k : |u'x| \leq c\} \quad \text{and} \quad H_{u,c}^{\text{out}} := \{x \in \mathbb{R}^k : |u'x| \geq c\}.$$

Henceforth, the superscript "in/out" is to be read as "in (resp., out)". We will further write $B(\theta, r)$ for the ball $\{x \in \mathbb{R}^k : \|x - \theta\| < r\}$ and $\bar{B}(\theta, r)$ for its closure.

Lemma A.2. Let P, Q be two probability measures over $\mathbb{R}^k$. Define

$$HD_{P,T}^{\text{in/out}}(\Sigma) := \inf_{u \in \mathcal{S}^{k-1}} P[T_P + \Sigma^{1/2}H_u^{\text{in/out}}],$$

with $H_u^{\text{in/out}} := H_{u,1}^{in/out}$. Then, for any $\Sigma \in \mathcal{P}_k$,

$$\begin{aligned} &|HD_{P,T}^{\text{in/out}}(\Sigma) - HD_{Q,T}^{\text{in/out}}(\Sigma)| \\ &\qquad \leq \sup_{C \in \mathcal{C}^{\text{in/out}}} |P[C] - Q[C]| + \sup_{C \in \mathcal{C}_0^{\text{in/out}}} |P[T_Q + C] - P[T_P + C]|, \end{aligned}$$

where we let $\mathcal{C}^{\text{in/out}} := \{\theta + H_{u,c}^{\text{in/out}} : (\theta, u, c) \in \mathbb{R}^k \times \mathcal{S}^{k-1} \times \mathbb{R}_0^+\}$ and $\mathcal{C}_0^{\text{in/out}} := \{H_{u,c}^{\text{in/out}} : (u, c) \in \mathcal{S}^{k-1} \times \mathbb{R}_0^+\}$.

Proof of Lemma A.2. We prove only the "in" result, since the proof of the "out" result is entirely similar. First assume that $HD_{P,T}^{\text{in}}(\Sigma) \geq HD_{Q,T}^{\text{in}}(\Sigma)$. Then, for any $\varepsilon > 0$, there exists $u_0 = u_0(\Sigma, Q, \varepsilon)$ such that $Q[T_Q + \Sigma^{1/2}H_{u_0}^{\text{in}}] \leq HD_{Q,T}^{\text{in}}(\Sigma) + \varepsilon$, so that

$$\begin{aligned} |HD_{P,T}^{\text{in}}(\Sigma) - HD_{Q,T}^{\text{in}}(\Sigma)| &= HD_{P,T}^{\text{in}}(\Sigma) - HD_{Q,T}^{\text{in}}(\Sigma) \\ &\leq P[T_P + \Sigma^{1/2}H_{u_0}^{\text{in}}] - Q[T_Q + \Sigma^{1/2}H_{u_0}^{\text{in}}] + \varepsilon \\ &\leq P[T_Q + \Sigma^{1/2}H_{u_0}^{\text{in}}] - Q[T_Q + \Sigma^{1/2}H_{u_0}^{\text{in}}] \\ &\qquad + P[T_P + \Sigma^{1/2}H_{u_0}^{\text{in}}] - P[T_Q + \Sigma^{1/2}H_{u_0}^{\text{in}}] + \varepsilon \\ &\leq \sup_{C \in \mathcal{C}^{\text{in}}} |P[C] - Q[C]| + \sup_{C \in \mathcal{C}_0^{\text{in}}} |P[T_Q + C] - P[T_P + C]| + \varepsilon. \end{aligned}$$

Similarly, if $HD_{P,T}^{\text{in}}(\Sigma) \leq HD_{Q,T}^{\text{in}}(\Sigma)$, then, for any $\varepsilon > 0$, there exists $u_1 = u_1(\Sigma, P, \varepsilon)$ such that $P[T_P + \Sigma^{1/2} H^{\mathrm{in}}_{u_1}] \leq HD^{\mathrm{in}}_{P,T}(\Sigma) + \varepsilon$, so that

$$\begin{aligned}
|HD^{\mathrm{in}}_{P,T}(\Sigma) - HD^{\mathrm{in}}_{Q,T}(\Sigma)| &= HD^{\mathrm{in}}_{Q,T}(\Sigma) - HD^{\mathrm{in}}_{P,T}(\Sigma) \\
&\leq Q[T_Q + \Sigma^{1/2} H^{\mathrm{in}}_{u_1}] - P[T_P + \Sigma^{1/2} H^{\mathrm{in}}_{u_1}] + \varepsilon \\
&\leq Q[T_Q + \Sigma^{1/2} H^{\mathrm{in}}_{u_1}] - P[T_Q + \Sigma^{1/2} H^{\mathrm{in}}_{u_1}] \\
&\qquad + P[T_Q + \Sigma^{1/2} H^{\mathrm{in}}_{u_1}] - P[T_P + \Sigma^{1/2} H^{\mathrm{in}}_{u_1}] + \varepsilon \\
&\leq \sup_{C \in \mathcal{C}^{\mathrm{in}}} |P[C] - Q[C]| + \sup_{C \in \mathcal{C}^{\mathrm{in}}_0} |P[T_Q + C] - P[T_P + C]| + \varepsilon.
\end{aligned}$$

Since, in both cases, the result holds for any $\varepsilon > 0$, the result is proved. □

Lemma A.3. Let P be a probability measure over $\mathbb{R}^k$ and K be a compact subset of $\mathbb{R}^k$. For any $c > 0$, let $s^K_P(c) := \sup_{(\theta,u) \in K \times \mathcal{S}^{k-1}} P[|u'(X-\theta)| \leq c]$ and write $s^K_P := s^K_P(0)$. Then (i) $s^K_P(c) \to s^K_P$ as $c \overset{>}{\to} 0$ and (ii) $s^K_P = P[u_0'(X-\theta_0) = 0]$ for some $(\theta_0, u_0) \in K \times \mathcal{S}^{k-1}$.

Proof of Lemma A.3. Clearly, $s^K_P(c)$ is increasing in c over $[0,\infty)$, which guarantees that $\tilde{s}^K_P := \lim_{c \overset{>}{\to} 0} s^K_P(c)$ exists and satisfies $\tilde{s}^K_P \geq s^K_P$. Now, fix an arbitrary decreasing sequence (c_n) converging to 0 and consider a sequence $((\theta_n, u_n))$ in $K \times \mathcal{S}^{k-1}$ such that

$$P[|u_n'(X - \theta_n) \leq c_n] \geq s^K_P(c_n) - (1/n).$$

Compactness of $K \times \mathcal{S}^{k-1}$ guarantees the existence of a subsequence $((\theta_{n_\ell}, u_{n_\ell}))$ that converges in $K \times \mathcal{S}^{k-1}$, to (θ_0, u_0) say. Clearly, without loss of generality, we can assume that $(u_{n_\ell}' u_0)$ is an increasing sequence and that $\|\theta_{n_\ell} - \theta_0\|$ is a decreasing sequence (if that is not the case, one can always extract a further subsequence which meets these monotonicity properties). Let then $\bar{B}_\ell := \bar{B}(\theta_0, \|\theta_{n_\ell} - \theta_0\|)$ and $C_\ell := \{u \in \mathcal{S}^{k-1} : u'u_0 \geq u_{n_\ell}' u_0\}$. Clearly, $\bar{B}_\ell$ and C_ℓ are decreasing sequences of sets, with $\cap_\ell \bar{B}_\ell = \{\theta_0\}$ and $\cap_\ell C_\ell = \{u_0\}$. Therefore,

$$\begin{aligned}
\lim_{\ell\to\infty} r_\ell &:= \lim_{\ell\to\infty} P[X \in \cup_{\theta \in \bar{B}_\ell} \cup_{u \in C_\ell} \{x = \theta + y : |u'y| \leq c_{n_\ell}\}] \\
&= P[u_0'(X - \theta_0) = 0].
\end{aligned}$$

Now, for any ℓ, $r_\ell \geq P[|u_{n_\ell}'(X - \theta_{n_\ell})| \leq c_{n_\ell}] \geq s^K_P(c_{n_\ell}) - (1/n_\ell)$, which implies that $s^K_P \geq P[|u_0'(X-\theta_0)| = 0] \geq \tilde{s}^K_P$. Therefore, $\tilde{s}^K_P = s^K_P = P[|u_0'(X-\theta_0)| = 0]$. □

Lemma A.4. Let P be a smooth probability measure over $\mathbb{R}^k$ and fix $\theta_0 \in \mathbb{R}^k$. Then

$$\sup_{(u,c) \in \mathcal{S}^{k-1} \times \mathbb{R}^+_0} |P[\theta + H^{\mathrm{in/out}}_{u,c}] - P[\theta_0 + H^{\mathrm{in/out}}_{u,c}]| \to 0$$

as $\theta \to \theta_0$.

Proof of Lemma A.4. We start with the "in" result. Fix $\varepsilon > 0$. Pick $c_1 > 0$ large enough to have $P[B(\theta_0, c_1/2)] \geq 1 - (\varepsilon/2)$. Pick then $c_0 \in (0, c_1)$ such that $s_P^{\bar{B}(\theta_0, 2c_1)}(c_0) < \varepsilon/2$ for any $c \in (0, c_0]$ (existence of such a c_0 is guaranteed by Lemma A.3 and the smoothness assumption on P). Then,

$$\sup_{(u,c)\in\mathcal{S}^{k-1}\times\mathbb{R}_0^+} |P[\theta + H_{u,c}^{\text{in}}] - P[\theta_0 + H_{u,c}^{\text{in}}]| = \max\big(Q_{0,c_0}^{\text{in}}, Q_{c_0,c_1}^{\text{in}}, Q_{c_1,\infty}^{\text{in}}\big),$$

where Q_{0,c_0}^{in}, Q_{c_0,c_1}^{in} and $Q_{c_1,\infty}^{\text{in}}$ are the suprema of $|P[\theta + H_{u,c}^{\text{in}}] - P[\theta_0 + H_{u,c}^{\text{in}}]|$ over $u \in \mathcal{S}^{k-1}$ and, respectively, $c \in (0, c_0]$, $c \in (c_0, c_1]$ and $c \in (c_1, \infty)$. Now, fix $\delta \in (0, \min(c_0, c_1/2))$ and let $\theta \in B(\theta_0, \delta)$.

(i) The choice of c_0 implies

$$\begin{aligned} Q_{0,c_0}^{\text{in}} &\leq \sup_{u\in\mathcal{S}^{k-1}} P[\theta + H_{u,c_0}^{\text{in}}] + \sup_{u\in\mathcal{S}^{k-1}} P[\theta_0 + H_{u,c_0}^{\text{in}}] \\ &\leq 2 \sup_{(\theta,u)\in\bar{B}(\theta_0,2c_1)\times\mathcal{S}^{k-1}} P[\theta + H_{u,c_0}^{\text{in}}] \leq \varepsilon. \end{aligned} \tag{A.4}$$

(ii) Let $u \in \mathcal{S}^{k-1}$ and $c \geq c_0$. Assume, without loss of generality, that $u'(\theta - \theta_0) \geq 0$ (the case $u'(\theta - \theta_0) \leq 0$ proceeds similarly). The set $\theta + H_{u,c}^{\text{in}}$ rewrites

$$\theta + H_{u,c}^{\text{in}} = \{\theta + x : |u'x| \leq c\} = \{\theta_0 + y : |u'y - (u'(\theta - \theta_0))| \leq c\}.$$

Therefore,

$$\begin{aligned} &|P[\theta + H_{u,c}^{\text{in}}] - P[\theta_0 + H_{u,c}^{\text{in}}]| \\ &\qquad \leq P\Big[\{\theta_0 + y : -c \leq u'y \leq -c + u'(\theta - \theta_0)\}\Big] \\ &\qquad\qquad + P\Big[\{\theta_0 + y : c \leq u'y \leq c + u'(\theta - \theta_0)\}\Big] \\ &\qquad \leq P\Big[\{\theta_0 + y : -c \leq u'y \leq -c + \delta\}\Big] \\ &\qquad\qquad + P\Big[\{\theta_0 + y : c \leq u'y \leq c + \delta\}\Big]. \end{aligned} \tag{A.5}$$

(iia) For $u \in \mathcal{S}^{k-1}$ and $c_0 < c \leq c_1$, set $\theta_1 = \theta_0 - cu + \delta u/2$ and $\theta_2 = \theta_0 + cu + \delta u/2$. It holds $\{\theta_0 + y : -c \leq u'y \leq -c + \delta\} = \{\theta_1 + x : |u'x| \leq \delta/2\}$ and $\{\theta_0 + y : c \leq u'y \leq c\delta\} = \{\theta_2 + x : |u'x| \leq \delta/2\}$. Since $s_P^{\bar{B}(\theta_0,2c_1)}(\delta/2) \leq s_P^{\bar{B}(\theta_0,2c_1)}(c_0) < \varepsilon/2$ and $\|\theta_\ell - \theta_0\| \leq c + \delta/2 < 2c_1$ ($\ell = 1, 2$), (A.5) yields

$$Q_{c_0,c_1}^{\text{in}} \leq 2 \sup_{(\theta,u)\in\bar{B}(\theta_0,2c_1)\times\mathcal{S}^{k-1}} P\Big[\{\theta + x : |u'x| \leq \delta/2\}\Big] \leq \varepsilon. \tag{A.6}$$

(iib) For $u \in \mathcal{S}^{k-1}$ and $c > c_1$, the sets $\{\theta_0 + y : -c \leq u'y \leq -c + \delta\}$ and $\{\theta_0 + y : c \leq u'y \leq c + \delta\}$ lie outside the ball $B(\theta_0, c_1/2)$ since $-c + \delta < -c + (c_1/2) < -c_1/2$ and $c > c_1 > c_1/2$, respectively. Therefore, it follows from (A.5) that

$$Q_{c_1,\infty}^{\text{in}} \leq 2P\big[\mathbb{R}^k \setminus B(\theta_0, c_1/2)\big] \leq \varepsilon. \tag{A.7}$$

According to (A.4), (A.6) and (A.7), all three quantities Q_{0,c_0}^{in}, Q_{c_0,c_1}^{in} and $Q_{c_1,\infty}^{\text{in}}$ are bounded by ε as soon as $\|\theta - \theta_0\| < \delta$, which concludes the proof of the "in" result.

The proof of the "out" result proceeds similarly. For the same choices of c_0, c_1 and δ, and the respective suprema Q^{out}_{0,c_0}, $Q^{\text{out}}_{c_0,c_1}$ and $Q^{\text{out}}_{c_1,\infty}$, it holds

$$\begin{aligned} Q^{\text{out}}_{c_1,\infty} &\leq \sup_{u\in\mathcal{S}^{k-1}} P[\theta + H^{\text{out}}_{u,c_1}] + \sup_{u\in\mathcal{S}^{k-1}} P[\theta_0 + H^{\text{out}}_{u,c_1}] \\ &\leq 2 \sup_{(\theta,u)\in\bar{B}(\theta_0,c_1/2)\times\mathcal{S}^{k-1}} P\big[\mathbb{R}^k \setminus B(\theta_0, c_1/2)\big] \leq \varepsilon. \end{aligned}$$

Moreover, the inequality $Q^{\text{out}}_{0,c_0} \leq \varepsilon$ follows from the fact that

$$|P[\theta + H^{\text{out}}_{u,c}] - P[\theta_0 + H^{\text{out}}_{u,c}]| \leq P[\theta + H^{\text{in}}_{u,c_0}] + P[\theta_0 + H^{\text{in}}_{u,c_0}]$$

for $c \leq c_0$. Finally, it can be proved that $Q^{\text{out}}_{c_0,c_1} \leq \varepsilon$ along the exact same lines as above. The "out" result follows. $\square$

Proof of Theorem 2.2. The collection $\mathcal{H}$ of all halfspaces in $\mathbb{R}^k$ is a Vapnik-Chervonenkis class; see, e.g., page 152 of Van der Vaart & Wellner (1996). Hence, Lemma 2.6.17 of the same implies that $\mathcal{H} \sqcap \mathcal{H} := \{H_1 \cap H_2 : H_1, H_2 \in \mathcal{H}\}$ and $\mathcal{H} \sqcup \mathcal{H} := \{H_1 \cup H_2 : H_1, H_2 \in \mathcal{H}\}$ are also Vapnik-Chervonenkis classes. Consequently, using henceforth the notation from Lemma A.2, $\mathcal{C}^{\text{in}}(\subset \mathcal{H} \sqcap \mathcal{H})$ and $\mathcal{C}^{\text{out}}(\subset \mathcal{H} \sqcup \mathcal{H})$ are themselves Vapnik-Chervonenkis classes, which implies that

$$\sup_{C\in\mathcal{C}^{\text{in}}} |P_n[C] - P[C]| \to 0 \quad \text{and} \quad \sup_{C\in\mathcal{C}^{\text{out}}} |P_n[C] - P[C]| \to 0 \tag{A.8}$$

almost surely as $n \to \infty$. Also, since $T_{P_n} \to T_P$ almost surely as $n \to \infty$, Lemma A.4 entails

$$\sup_{C\in\mathcal{C}_0^{\text{in}}} |P[T_{P_n} + C] - P[T_P + C]| \to 0 \quad \text{and} \quad \sup_{C\in\mathcal{C}_0^{\text{out}}} |P[T_{P_n} + C] - P[T_P + C]| \to 0 \tag{A.9}$$

almost surely as $n \to \infty$.

Now, by using Lemma A.2, we obtain that, for any $\Sigma \in \mathcal{P}_k$,

$$\begin{aligned} &|HD^{\text{sc}}_{P_n,T}(\Sigma) - HD^{\text{sc}}_{P,T}(\Sigma)| \\ &\quad = |\min(HD^{\text{in}}_{P_n,T}(\Sigma), HD^{\text{out}}_{P_n,T}(\Sigma)) - \min(HD^{\text{in}}_{P,T}(\Sigma), HD^{\text{out}}_{P,T}(\Sigma))| \\ &\quad \leq \max\big(|HD^{\text{in}}_{P_n,T}(\Sigma) - HD^{\text{in}}_{P,T}(\Sigma)|, |HD^{\text{out}}_{P_n,T}(\Sigma) - HD^{\text{out}}_{P,T}(\Sigma)|\big) \\ &\quad \leq \max\Big(\sup_{C\in\mathcal{C}^{\text{in}}} |P_n[C] - P[C]| + \sup_{C\in\mathcal{C}_0^{\text{in}}} |P[T_{P_n} + C] - P[T_P + C]|, \\ &\qquad\qquad \sup_{C\in\mathcal{C}^{\text{out}}} |P_n[C] - P[C]| + \sup_{C\in\mathcal{C}_0^{\text{out}}} |P[T_{P_n} + C] - P[T_P + C]|\Big). \end{aligned}$$

Consequently,

$$\begin{aligned}
&\sup_{\Sigma\in\mathcal{P}_k} |HD^{\mathrm{sc}}_{P_n,T}(\Sigma) - HD^{\mathrm{sc}}_{P,T}(\Sigma)| \\
&\qquad\qquad \leq \max\Big(\sup_{C\in\mathcal{C}^{\mathrm{in}}} |P_n[C]-P[C]| + \sup_{C\in\mathcal{C}^{\mathrm{in}}_0} |P[T_{P_n}+C]-P[T_P+C]|, \\
&\qquad\qquad\qquad\quad \sup_{C\in\mathcal{C}^{\mathrm{out}}} |P_n[C]-P[C]| + \sup_{C\in\mathcal{C}^{\mathrm{out}}_0} |P[T_{P_n}+C]-P[T_P+C]|\Big),
\end{aligned}$$

which, in view of (A.8) and (A.9), establishes the result. □

We close this section by proving that the Tukey median θ_P is strongly consistent without any assumption on P.

Lemma A.5. Let P be a probability measure over $\mathbb{R}^k$ and P_n denote the empirical measure associated with a random sample of size n from P. Then $\theta_{P_n} \to \theta_P$ almost surely as $n\to\infty$.

Proof of Lemma A.5. For any $\theta\in\mathbb{R}^k$ and any probability measure Q over $\mathbb{R}^k$, denote by $HD^{\mathrm{loc}}_Q(\theta)$ the location halfspace depth of θ with respect to Q. Recall that we defined θ_Q as the barycentre of $M^{\mathrm{loc}}_Q=\{\theta\in\mathbb{R}^k : HD^{\mathrm{loc}}_Q(\theta)=\max_{\eta\in\mathbb{R}^k} HD^{\mathrm{loc}}_Q(\eta)\}$. It is well known that $\theta\mapsto HD^{\mathrm{loc}}_Q(\theta)$ is upper semicontinuous; see, e.g., Lemma 6.1 in Donoho & Gasko (1992). In general, this function is not uniquely maximised at θ_Q. However, it is easy to define a modified depth function $\theta\mapsto HD^{\mathrm{loc}}_{Q,\mathrm{mod}}(\theta)$ that is still upper semicontinuous, agrees with $\theta\mapsto HD^{\mathrm{loc}}_Q(\theta)$ on $\mathbb{R}^k/M^{\mathrm{loc}}_Q$, and for which θ_Q is the unique maximiser. In view of the uniform consistency of location halfspace depth (see, e.g., (6.2) and (6.6) in Donoho & Gasko, 1992), the result then follows from Theorem 2.12 and Lemma 14.3 in Kosorok (2008). □

A.2 Proofs from Section 3

Proof of Theorem 3.1. (i) Fix $u\in\mathcal{S}^{k-1}$. Since $H^{\mathrm{in}}_u := \{x\in\mathbb{R}^k : |u'x|\leq 1\}$ is a closed subset of $\mathbb{R}^k$, the mapping $P\mapsto P[H^{\mathrm{in}}_u]$ is upper semicontinuous for weak convergence. Now, Slutzky's lemma entails that, as $d_F(\Sigma,\Sigma_0)\to 0$, the measure defined by $B\mapsto P[T_P+\Sigma^{1/2}B]$ converges weakly to the one defined by $B\mapsto P[T_P+\Sigma_0^{1/2}B]$. Therefore, $\Sigma\mapsto P[T_P+\Sigma^{1/2}H^{\mathrm{in}}_u]$ is upper F-semicontinuous at Σ_0. Since $H^{\mathrm{out}}_u := \{x\in\mathbb{R}^k : |u'x|\geq 1\}$ is also a closed subset of $\mathbb{R}^k$, the same argument shows that $\Sigma\mapsto P[T_P+\Sigma^{1/2}H^{\mathrm{out}}_u]$ is upper F-semicontinuous at Σ_0. Therefore

$$\Sigma\mapsto HD^{\mathrm{sc}}_{P,T}(\Sigma)=\min\Big(\inf_{u\in\mathcal{S}^{k-1}} P[T_P+\Sigma^{1/2}H^{\mathrm{in}}_u],\ \inf_{u\in\mathcal{S}^{k-1}} P[T_P+\Sigma^{1/2}H^{\mathrm{out}}_u]\Big),$$

is upper F-semicontinuous (recall that the infimum of a collection of upper semicontinuous functions is upper semicontinuous).

(ii) The result directly follows from the fact that $R^{\mathrm{sc}}_{P,T}(\alpha)$ is the inverse image of $[\alpha,+\infty)$ by the upper F-semicontinuous function $\Sigma\mapsto HD^{\mathrm{sc}}_{P,T}(\Sigma)$.

(iii) Fix a sequence (Σ_n) in $\mathcal{P}_k$ converging to Σ_0 with respect to the Frobenius distance.

With the same notation as in the proof of (i), note that, for any Σ,

$$HD^{\rm sc}_{P,T}(\Sigma) \;=\; \inf_{u\in\mathcal{S}^{k-1}} \min\Big(P[T_P+\Sigma^{1/2}H^{\rm in}_u],P[T_P+\Sigma^{1/2}H^{\rm out}_u]\Big).$$

For any n, pick then $u_n(\in\mathcal{S}^{k-1})$ such that

$$\min\Big(P[T_P+\Sigma_n^{1/2}H^{\rm in}_{u_n}],P[T_P+\Sigma_n^{1/2}H^{\rm out}_{u_n}]\Big)\leq HD^{\rm sc}_{P,T}(\Sigma_n)+\frac{1}{n}\cdot$$

Compactness of $\mathcal{S}^{k-1}$ implies that we can extract a subsequence (u_{n_ℓ}) of (u_n) that converges to $u_0(\in\mathcal{S}^{k-1})$. Writing $\mathbb{I}[C]$ for the indicator function of the set C, the dominated convergence theorem then yields that

$$\begin{aligned}P[T_P+\Sigma_{n_\ell}^{1/2}H^{\rm in}_{u_{n_\ell}}]&-P[T_P+\Sigma_0^{1/2}H^{\rm in}_{u_0}]\\&=\int_{\mathbb{R}^k}(\mathbb{I}[T_P+\Sigma_{n_\ell}^{1/2}H^{\rm in}_{u_{n_\ell}}]-\mathbb{I}[T_P+\Sigma_0^{1/2}H^{\rm in}_{u_0}])\,dP\to 0\end{aligned}$$

as $\ell\to\infty$ (the smoothness assumption on P guarantees that $\mathbb{I}[T_P+\Sigma_{n_\ell}^{1/2}H^{\rm in}_{u_{n_\ell}}]-\mathbb{I}[T_P+\Sigma_0^{1/2}H^{\rm in}_{u_0}]\to 0$ P-almost everywhere). Proceeding in the same way, we obtain that $P[T_P+\Sigma_{n_\ell}^{1/2}H^{\rm out}_{u_{n_\ell}}]-P[T_P+\Sigma_0^{1/2}H^{\rm out}_{u_0}]\to 0$ as $\ell\to\infty$. Consequently,

$$\begin{aligned}\liminf_{n\to\infty}HD^{\rm sc}_{P,T}(\Sigma_n)\;&=\;\liminf_{n\to\infty}\min\Big(P[T_P+\Sigma_n^{1/2}H^{\rm in}_{u_n}],P[T_P+\Sigma_n^{1/2}H^{\rm out}_{u_n}]\Big)\\&=\;\liminf_{\ell\to\infty}\min\Big(P[T_P+\Sigma_{n_\ell}^{1/2}H^{\rm in}_{u_{n_\ell}}],P[T_P+\Sigma_{n_\ell}^{1/2}H^{\rm out}_{u_{n_\ell}}]\Big)\\&=\;\min\Big(P[T_P+\Sigma_0^{1/2}H^{\rm in}_{u_0}],P[T_P+\Sigma_0^{1/2}H^{\rm out}_{u_0}]\Big)\\&\geq\;HD^{\rm sc}_{P,T}(\Sigma_0).\end{aligned}$$

We conclude that, if P is smooth at T_p, then $\Sigma\to HD^{\rm sc}_{P,T}(\Sigma)$ is also lower F-semicontinuous, hence F-continuous. $\square$

Proof of Theorem 3.2. Fix $\alpha>0$. Note that $\lambda_1(\Sigma)\geq\|\Sigma\|_F/\sqrt{k}\geq(\|\Sigma-I_k\|_F-\|I_k\|_F)/\sqrt{k}$ for any $\Sigma\in\mathcal{P}_k$. Therefore, denoting by $v_1(\Sigma)$ an arbitrary unit eigenvector associated with $\lambda_1(\Sigma)$, we have that, for any $\Sigma\notin B_F(I_k,r)$,

$$\begin{aligned}HD^{\rm sc}_{P,T}(\Sigma)&\leq\inf_{u\in\mathcal{S}^{k-1}}P\big[|u'(X-T_P)|\geq\sqrt{u'\Sigma u}\,\big]\\&\leq P[|v_1'(\Sigma)(X-T_P)|\geq\sqrt{\lambda_1(\Sigma)}]\leq P\Big[\|X-T_P\|\geq\tfrac{(r-1)^{1/2}}{k^{1/4}}\Big],\end{aligned}$$

which can be made strictly smaller than α for r large enough. This confirms that, for r large enough, $R^{\rm sc}_{P,T}(\alpha)$ is included in the ball $B_F(I_k,r)$, hence is F-bounded. $\square$

Proof of Theorem 3.3. (i) With $\Sigma_t=(1-t)\Sigma_a+t\Sigma_b$, we clearly have that, for any $u\in\mathcal{S}^{k-1}$,

$\min(u'\Sigma_a u, u'\Sigma_b u) \le u'\Sigma_t u \le \max(u'\Sigma_a u, u'\Sigma_b u)$. This entails that, for any $u \in \mathcal{S}^{k-1}$,

$$\begin{aligned}
&P\big[|u'(X - T_P)| \le \sqrt{u'\Sigma_t u}\,\big] \\
&\qquad\qquad \ge \min(P\big[|u'(X - T_P)| \le \sqrt{u'\Sigma_a u}\,\big], P\big[|u'(X - T_P)| \le \sqrt{u'\Sigma_b u}\,\big]) \\
&\qquad\qquad \ge \min(HD^{\rm sc}_{P,T}(\Sigma_a), HD^{\rm sc}_{P,T}(\Sigma_b))
\end{aligned}$$

and

$$\begin{aligned}
&P\big[|u'(X - T_P)| \ge \sqrt{u'\Sigma_t u}\,\big] \\
&\qquad\qquad \ge \min(P\big[|u'(X - T_P)| \ge \sqrt{u'\Sigma_a u}\,\big], P\big[|u'(X - T_P)| \ge \sqrt{u'\Sigma_b u}\,\big]) \\
&\qquad\qquad \ge \min(HD^{\rm sc}_{P,T}(\Sigma_a), HD^{\rm sc}_{P,T}(\Sigma_b)).
\end{aligned}$$

The result follows. (ii) If both $\Sigma_a, \Sigma_b \in R^{\rm sc}_{P,T}(\alpha)$, then Part (i) of the result entails that, for any $t \in [0, 1]$, $HD^{\rm sc}_{P,T}(\Sigma_t) \ge \min(HD^{\rm sc}_{P,T}(\Sigma_a), HD^{\rm sc}_{P,T}(\Sigma_b)) \ge \alpha$, so that $\Sigma_t \in R^{\rm sc}_{P,T}(\alpha)$. □

A.3 Proofs from Section 4

For the sake of completeness, we prove the following result.

Lemma A.6. Let R be a g-bounded subset of $\mathcal{P}_k$. Then R is totally g-bounded, that is, for any ε, there exist Σ_i, $i = 1, \ldots, m = m(\varepsilon)$ such that $R \subset \cup_{i=1}^m B_g(\Sigma_i, \varepsilon)$.

Proof of Lemma A.6. As a mapping from the metric space $(\mathcal{S}_k, d_F)$ (recall that d_F denotes the Frobenius distance) to the metric space $(\mathcal{P}_k, d_g)$, $A \mapsto \exp(A)$ is continuous; see the proof of Proposition 10 in Bhatia & Holbrook (2006). Denoting, for any $A \in \mathcal{S}_k$, as $\mathrm{vech}(A)$ the vector obtained by stacking the upper-diagonal entries of A on top of each other, the mapping $v \mapsto \mathrm{vech}^{-1}(v)$ from $(\mathbb{R}^{k(k+1)/2}, d_E)$ (equipped with the usual Euclidean distance d_E) to $(\mathcal{S}_k, d_F)$ is trivially continuous, so that the mapping $f : (\mathbb{R}^{k(k+1)/2}, d_E) \to (\mathcal{P}_k, d_g) : v \mapsto f(v) := \exp(\mathrm{vech}^{-1}(v))$ is also continuous.

Now, fix $\varepsilon > 0$, pick $r > 0$ such that R is included in the closed ball $\bar{B} := \bar{B}_g(I_k, r) := \{\Sigma \in \mathcal{P}_k : d_g(\Sigma, I_k) \le r\}$, and consider the resulting open covering $\{B_g(\Sigma, \varepsilon) : \Sigma \in \bar{B}\}$ of $\bar{B}$. From continuity, $\mathcal{C} := \{f^{-1}(B_g(\Sigma, \varepsilon)) : \Sigma \in \bar{B}\}$ is an open covering of the closed set $f^{-1}(\bar{B})$ in $\mathbb{R}^{k(k+1)/2}$. It is easy to check that, for any $\Sigma \in \bar{B}$, $\lambda_1(\Sigma) \le \exp(r/\sqrt{k})$, so that $f^{-1}(\bar{B})$ is bounded, hence compact. Therefore, a finite subcovering $\{f^{-1}(B_g(\Sigma_i, \varepsilon)) : i = 1, \ldots, m\}$ of $f^{-1}(\bar{B})$ can be extracted from $\mathcal{C}$, which provides the desired finite covering $\{B_g(\Sigma_i, \varepsilon) : i = 1, \ldots, m\}$ of $\bar{B}$, hence of R, with open g-balls of radius ε. □

Proof of Theorem 4.2. Assume first that $s_{P,T} < 1/2$ and fix $\varepsilon > 0$. We will then prove that $R^{\rm sc}_{P,T}(s_{P,T} + \varepsilon)$ is g-bounded by showing that, for $r > 0$ large enough, it is included in the

g-ball $B_g(I_k, r)$. To do so, first note that (4.1) entails

$$d_g(\Sigma, I_k) = \sqrt{\sum_{i=1}^{k} (\log \lambda_i(\Sigma))^2} \leq \sqrt{k}\, \max(|\log \lambda_1(\Sigma)|, |\log \lambda_k(\Sigma)|)$$
$$= \sqrt{k}\, \max(\log \lambda_1(\Sigma), \log \lambda_k^{-1}(\Sigma)). \tag{A.10}$$

Therefore, $\Sigma \notin B_g(I_k, r)$ implies that (i) $\lambda_1(\Sigma) > \exp(r/\sqrt{k})$ or (ii) $\lambda_k(\Sigma) < \exp(-r/\sqrt{k})$ (or both). In case (i),

$$\begin{aligned} HD^{\mathrm{sc}}_{P,T}(\Sigma) &\leq \inf_{u \in \mathcal{S}^{k-1}} P\big[|u'(X - T_P)| \geq \sqrt{u'\Sigma u}\,\big] \\ &\leq P[|v_1'(\Sigma)(X - T_P)| \geq \sqrt{\lambda_1(\Sigma)}] \\ &\leq P[|v_1'(\Sigma)(X - T_P)| \geq \exp(r/2\sqrt{k})] \\ &\leq P[\|X - T_P\| \geq \exp(r/2\sqrt{k})], \end{aligned}$$

which can be made smaller than ε (hence, smaller than $s_{P,T}+\varepsilon$) for r large enough. In case (ii), we have that, using the notation $s_P^K(\cdot)$ from Lemma A.3,

$$\begin{aligned} HD^{\mathrm{sc}}_{P,T}(\Sigma) &\leq \inf_{u \in \mathcal{S}^{k-1}} P\big[|u'(X - T_P)| \leq \sqrt{u'\Sigma u}\,\big] \\ &\leq P\big[|v_k'(\Sigma)(X - T_P)| \leq \lambda_k^{1/2}(\Sigma)\big] \\ &\leq s_P^{\{T_P\}}(\lambda_k^{1/2}(\Sigma)) \leq s_P^{\{T_P\}}(\exp(-r/2\sqrt{k})), \end{aligned}$$

which, in view of Lemma A.3(i), can be made smaller than $s_{P,T} + \varepsilon$ for r large enough. We conclude that, for $\alpha > s_{P,T}$, $R^{\mathrm{sc}}_{P,T}(\alpha)$ is g-bounded, hence also (Lemma A.6) totally g-bounded. Since it is also g-closed (which follows from Theorem 4.1(ii)), it is g-compact (recall from Section 4 that, in a complete metric space, any closed and totally bounded set is compact).

Finally, if $s_{P,T} \geq 1/2$, then, with $u_0 \in \mathcal{S}^{k-1}$ such that $P[|u_0'(X - T_P)| = 0] = s_{P,T}$ (existence is guaranteed in Lemma A.3(ii); take $K = \{T_P\}$ there), we have $HD^{\mathrm{sc}}_{P,T}(\Sigma) \leq P[|u_0'(X - T_P)| \geq \sqrt{u_0'\Sigma u_0}] \leq P[|u_0'(X - T_P)| > 0] = 1 - s_{P,T}$, so that $R^{\mathrm{sc}}_{P,T}(\alpha)$ is empty for any $\alpha > 1 - s_{P,T} = \alpha_{P,T}$. □

Proof of Theorem 4.3. By assumption, $\alpha_{*P,T} = \sup_{\Sigma \in \mathcal{P}_k} HD^{\mathrm{sc}}_{P,T}(\Sigma) \geq \alpha_{P,T}$ since $R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$ is non-empty. If $\alpha_{*P,T} = \alpha_{P,T}$, then the result holds since the maximal depth $\alpha_{*P,T}$ is achieved at any scatter matrix in the non-empty set $R^{\mathrm{sc}}_{P,T}(\alpha_{P,T})$. Assume then that $\alpha_{*P,T} > \alpha_{P,T}$. Fix $\delta > 0$ such that $\alpha_{*P,T} - \delta > \alpha_{P,T}$, so that $R^{\mathrm{sc}}_{P,T}(\alpha_{*P,T} - \delta)$ is g-compact (Theorem 4.2). For any positive integer n, it is possible to pick a scatter matrix Σ_n with $HD^{\mathrm{sc}}_{P,T}(\Sigma_n) \geq \alpha_{*P,T} - (\delta/n)$. The g-compactness of $R^{\mathrm{sc}}_{P,T}(\alpha_{*P,T} - \delta)$ implies that there exists a subsequence (Σ_{n_ℓ}) that g-converges in $R^{\mathrm{sc}}_{P,T}(\alpha_{*P,T} - \delta)$, to Σ_* say. For any $\varepsilon \in (0, \delta)$, all terms of (Σ_{n_ℓ}) are eventually in the g-closed set $R^{\mathrm{sc}}_{P,T}(\alpha_{*P,T} - \varepsilon)$, so that its g-limit Σ_* must also belong to $R^{\mathrm{sc}}_{P,T}(\alpha_{*P,T} - \varepsilon)$. For any such ε, we thus have $\alpha_{*P,T} - \varepsilon \leq HD^{\mathrm{sc}}_{P,T}(\Sigma_*) \leq \alpha_{*P,T}$, which proves that $HD^{\mathrm{sc}}_{P,T}(\Sigma_*) = \alpha_{*P,T}$. □

The proof of Theorem 4.4 requires the following preliminary result.

Lemma A.7. Fix $\Sigma \in \mathcal{P}_k$ with $\max(\text{diag}(\Sigma)) \leq 1$. Then $\max_s s'\Sigma^{-1}s \geq k$ (where $\max_s$ is the maximum over $s = (s_1, \ldots, s_k) \in \{-1, 1\}^k$), with equality if and only if $\Sigma = I_k$.

Proof of Lemma A.7. We prove the result by induction. Clearly, the result holds for $k = 1$. Assume then that the result holds for k. Writing

$$\Sigma = \begin{pmatrix} \Sigma_- & v \\ v' & \Sigma_{k+1,k+1} \end{pmatrix} \quad \text{and} \quad s = \begin{pmatrix} s_- \\ s_{k+1} \end{pmatrix},$$

the classical formula for the inverse of a block partitioned matrix yields

$$\begin{aligned} s'\Sigma^{-1}s &= s'_-\Sigma_-^{-1}s_- + \frac{(s'_-\Sigma_-^{-1}v - s_{k+1})^2}{\Sigma_{k+1,k+1} - v'\Sigma_-^{-1}v} \\ &= s'_-\Sigma_-^{-1}s_- + 1 + \frac{(s'_-\Sigma_-^{-1}v - s_{k+1})^2 - (\Sigma_{k+1,k+1} - v'\Sigma_-^{-1}v)}{\Sigma_{k+1,k+1} - v'\Sigma_-^{-1}v}\cdot \end{aligned} \tag{A.11}$$

By induction assumption, there exists s_- such that

$$s'\Sigma^{-1}s \geq k + 1 + \frac{(s'_-\Sigma_-^{-1}v - s_{k+1})^2 - (\Sigma_{k+1,k+1} - v'\Sigma_-^{-1}v)}{\Sigma_{k+1,k+1} - v'\Sigma_-^{-1}v}\cdot \tag{A.12}$$

Now, irrespective of s_-, choosing $s_{k+1} = -\text{sign}(s'_-\Sigma_-^{-1}v)$ yields

$$\begin{aligned} &(s'_-\Sigma_-^{-1}v - s_{k+1})^2 - (\Sigma_{k+1,k+1} - v'\Sigma_-^{-1}v) \\ &\qquad = (s'_-\Sigma_-^{-1}v)^2 + 2|s'_-\Sigma_-^{-1}v| + v'\Sigma_-^{-1}v + 1 - \Sigma_{k+1,k+1} \geq 0, \end{aligned}$$

since $\Sigma_{k+1,k+1} \leq 1$. Jointly with (A.12), this provides $\max_s s'\Sigma^{-1}s \geq k + 1$.

Now, assume that $\max_s s'\Sigma^{-1}s = k + 1$. We consider two cases. (a) $\max_{s_-} s'_-\Sigma_-^{-1}s_- > k$. Pick an arbitrary s_{*-} such that $s'_{*-}\Sigma_-^{-1}s_{*-} > k$. Then, with $s_* = (s'_{*-}, s_{*,k+1})'$, we have

$$s'_*\Sigma^{-1}s_* > k + 1 + \frac{(s'_{*-}\Sigma_-^{-1}v - s_{*,k+1})^2 - (\Sigma_{k+1,k+1} - v'\Sigma_-^{-1}v)}{\Sigma_{k+1,k+1} - v'\Sigma_-^{-1}v}\cdot$$

Choosing again $s_{*,k+1} = -\text{sign}(s'_{*-}\Sigma_-^{-1}v)$ makes the third term of the righthand side non-negative, which implies that $\max_s s'\Sigma^{-1}s > k + 1$, a contradiction. (b) $\max_{s_-} s'_-\Sigma_-^{-1}s_- = k$. By induction assumption, we must then have $\Sigma_- = I_k$. For any $s = (s'_-, s_{k+1})'$, (A.11) thus yields

$$s'\Sigma^{-1}s = s'_-s_- + \frac{(s'_-v - s_{k+1})^2}{\Sigma_{k+1,k+1} - v'v} = k + \frac{(s'_-v - s_{k+1})^2}{\Sigma_{k+1,k+1} - v'v}\cdot$$

Since $\max_s s'\Sigma^{-1}s = k + 1$, we must have that

$$1 = \max_s \frac{(s'_-v - s_{k+1})^2}{\Sigma_{k+1,k+1} - v'v} = \frac{(1 + \sum_{\ell=1}^k |v_\ell|)^2}{\Sigma_{k+1,k+1} - v'v} =: \frac{c}{d}\cdot$$

Since $c \geq 1$ and $d \leq 1$ (recall that $\Sigma_{k+1,k+1} \leq 1$), this imposes that $c = d = 1$, which leads to $v = 0$ and $\Sigma_{k+1,k+1} = 1$. Jointly with $\Sigma_- = I_k$, this shows that we must have $\Sigma = I_{k+1}$, which establishes the result. $\square$

Proof of Theorem 4.4. First note that, with $\Sigma_* = \sqrt{k} I_k$, (2.7) yields

$$HD^{\mathrm{sc}}_{P,T}(\Sigma_*) = 2\min\Big(\Psi\big(k^{-1/4}\big) - \tfrac{1}{2}, 1 - \Psi\big(k^{1/4}\big)\Big) = \tfrac{2}{\pi}\arctan\big(k^{-1/4}\big)$$

and fix an arbitrary $\Sigma \in \mathcal{P}_k$. If $\max(\mathrm{diag}(\Sigma)) > \sqrt{k}$, then

$$HD^{\mathrm{sc}}_{P,T}(\Sigma) \leq 2\big(1 - \Psi\big(\sqrt{\max(\mathrm{diag}(\Sigma))}\big)\big) < 2\big(1 - \Psi\big(k^{1/4}\big)\big) = HD^{\mathrm{sc}}_{P,T}(\Sigma_*). \tag{A.13}$$

If $\max(\mathrm{diag}(\Sigma)) \leq \sqrt{k}$, then Lemma A.7 yields

$$\max_s s'\Sigma^{-1}s = k^{-1/2}\max_s s'(k^{-1/2}\Sigma)^{-1}s \geq k^{1/2},$$

so that

$$\begin{aligned} HD^{\mathrm{sc}}_{P,T}(\Sigma) &\leq 2\big(\Psi\big(1/\max_s \sqrt{s'\Sigma^{-1}s}\,\big) - \tfrac{1}{2}\big) \\ &\leq 2\big(\Psi\big(k^{-1/4}\big) - \tfrac{1}{2}\big) = HD^{\mathrm{sc}}_{P,T}(\Sigma_*). \end{aligned} \tag{A.14}$$

We conclude that $HD^{\mathrm{sc}}_{P,T}(\Sigma_*) \geq HD^{\mathrm{sc}}_{P,T}(\Sigma)$ for any $\Sigma \in \mathcal{P}_k$. Now, assume that $HD^{\mathrm{sc}}_{P,T}(\Sigma) = HD^{\mathrm{sc}}_{P,T}(\Sigma_*)$ for some $\Sigma \in \mathcal{P}_k$. If $\max(\mathrm{diag}(\Sigma)) > \sqrt{k}$, we can only have $HD^{\mathrm{sc}}_{P,T}(\Sigma) < HD^{\mathrm{sc}}_{P,T}(\Sigma_*)$, as showed in (A.13). Thus we must have $\max(\mathrm{diag}(\Sigma)) \leq \sqrt{k}$, and by assumption, all inequalities in (A.14) should be equalities. In view of Lemma A.7, this implies that $k^{-1/2}\Sigma = I_k$, which establishes the result. $\square$

A.4 Proofs from Sections 5 and 6

Proof of Theorem 5.1. We start with the case $\theta_0 = 0$ and $\Sigma_0 = I_k$, for which

$$\begin{aligned} &\min(P\big[|u'X| \leq \sqrt{u'\Sigma u}\,\big], P\big[|u'X| \geq \sqrt{u'\Sigma u}\,\big]) \\ &\qquad = \min(P[|X_1| \leq \sqrt{u'\Sigma u}], P[|X_1| \geq \sqrt{u'\Sigma u}]) \end{aligned}$$

for any $u \in \mathcal{S}^{k-1}$, so that (note that the affine equivariance of T_P entails that $T_P = 0$)

$$\begin{aligned} HD^{\mathrm{sc}}_{P,T}(\Sigma) &= \inf_{z \in \mathrm{Sp}(\Sigma)} \min(P[X_1^2 \leq z], P[X_1^2 \geq z]) \\ &\leq \min(P[X_1^2 \leq 1], P[X_1^2 \geq 1]) = HD^{\mathrm{sc}}_{P,T}(I_k), \end{aligned}$$

where the equality holds if and only if $\mathrm{Sp}(\Sigma) \subset \mathcal{I}_{\mathrm{MSD}}[X_1]$. The result for a general location θ_0 and scatter Σ_0 readily follows from affine invariance and the identity $\mathrm{Sp}(AB) = \mathrm{Sp}(BA)$. (ii) By definition, if $\mathcal{I}_{\mathrm{MSD}}[Z_1]$ is a singleton, then this singleton must be $\{1\}$. Consequently, if $HD^{\mathrm{sc}}_{P,T}(\Sigma) = HD^{\mathrm{sc}}_{P,T}(\Sigma_0)$, then Part (i) of the result entails that $\lambda_k(\Sigma_0^{-1}\Sigma) = \lambda_1(\Sigma_0^{-1}\Sigma) = 1$. This implies that $\Sigma_0^{-1}\Sigma = I_k$, which establishes the result. $\square$

We turn to the proofs of Theorems 5.2 and 6.1, that require the following preliminary results for the elliptical case (Lemma A.8) and for the independent Cauchy case (Lemma A.9).

Lemma A.8. For any $\Sigma_a, \Sigma_b \in \mathcal{P}_k$ and $t \in [0,1]$, let $\tilde{\Sigma}_t := \Sigma_a^{1/2}\big(\Sigma_a^{-1/2}\Sigma_b \Sigma_a^{-1/2}\big)^t\Sigma_a^{1/2}$ and $\Sigma_t^* := ((1-t)\Sigma_a^{-1}+t\Sigma_b^{-1})^{-1}$. Then (i) $\lambda_1(\Sigma_t^*) \leq \lambda_1(\tilde{\Sigma}_t) \leq \max(\lambda_1(\Sigma_a), \lambda_1(\Sigma_b))$ and (ii) $\lambda_k(\tilde{\Sigma}_t) \geq \lambda_k(\Sigma_t^*) \geq \min(\lambda_k(\Sigma_a), \lambda_k(\Sigma_b))$.

Proof of Lemma A.8. (i) With the usual order on positive semidefinite matrices ($A \leq B$ iff $B - A$ is positive semidefinite), the (weighted) harmonic-geometric-arithmetic inequality (see, e.g., Lemma 2.1(vii) in Lawson & Lim, 2013)

$$\Sigma_t^* \leq \tilde{\Sigma}_t \leq \Sigma_t := (1-t)\Sigma_a + t\Sigma_b \tag{A.15}$$

holds for any $t \in [0,1]$. This implies that

$$\lambda_1(\Sigma_t^*) \leq \lambda_1(\tilde{\Sigma}_t) \leq \lambda_1(\Sigma_t). \tag{A.16}$$

Indeed, if, e.g., the second inequality in (A.16) does not hold (the argument for the first inequality is strictly the same), then, denoting as $\tilde{v}_{1t}$ an arbitrary eigenvector associated with $\lambda_1(\tilde{\Sigma}_t)$, we have $\tilde{v}_{1t}'\tilde{\Sigma}_t\tilde{v}_{1t} = \lambda_1(\tilde{\Sigma}_t) > \lambda_1(\Sigma_t) \geq \tilde{v}_{1t}'\Sigma_t\tilde{v}_{1t}$, which contradicts (A.15). Hence, (A.16) holds and provides

$$\begin{aligned}
\lambda_1(\Sigma_t^*) \leq \lambda_1(\tilde{\Sigma}_t) &\leq \max_{u\in\mathcal{S}^{d-1}} u'((1-t)\Sigma_a + t\Sigma_b)u \\
&\leq (1-t)\max_{u\in\mathcal{S}^{d-1}} u'\Sigma_a u + t\max_{u\in\mathcal{S}^{d-1}} u'\Sigma_b u \\
&= (1-t)\lambda_1(\Sigma_a) + t\lambda_1(\Sigma_b) \\
&\leq \max(\lambda_1(\Sigma_a), \lambda_1(\Sigma_b)),
\end{aligned}$$

as was to be showed. (ii) Proceeding in a similar way as above, it is readily showed that (A.15) implies that $\lambda_k(\tilde{\Sigma}_t) \geq \lambda_k(\Sigma_t^*)$. Using this, we obtain

$$\begin{aligned}
&\lambda_k(\tilde{\Sigma}_t) \geq \lambda_k(\Sigma_t^*) = \lambda_1^{-1}((\Sigma_t^*)^{-1}) \\
&= \Big(\max_{u\in\mathcal{S}^{d-1}} u'((1-t)\Sigma_a^{-1} + t\Sigma_b^{-1})u\Big)^{-1} \geq \Big((1-t)\lambda_1(\Sigma_a^{-1}) + t\lambda_1(\Sigma_b^{-1})\Big)^{-1} \\
&= \Big((1-t)\lambda_k^{-1}(\Sigma_a) + t\lambda_k^{-1}(\Sigma_b)\Big)^{-1} \geq \min(\lambda_k(\Sigma_a), \lambda_k(\Sigma_b)),
\end{aligned}$$

since any weighted harmonic mean of two real numbers is a convex linear combination of these. □

Lemma A.9. For any $\Sigma_a, \Sigma_b \in \mathcal{P}_k$ and $t \in [0,1]$, let $\tilde{\Sigma}_t := \Sigma_a^{1/2}\big(\Sigma_a^{-1/2}\Sigma_b \Sigma_a^{-1/2}\big)^t\Sigma_a^{1/2}$ and $\Sigma_t^* := ((1-t)\Sigma_a^{-1} + t\Sigma_b^{-1})^{-1}$. Then,

$$\max(\mathrm{diag}(\Sigma_t^*)) \leq \max(\mathrm{diag}(\tilde{\Sigma}_t)) \leq \max(\max(\mathrm{diag}(\Sigma_a)), \max(\mathrm{diag}(\Sigma_b))) \tag{A.17}$$

and

$$\max_s s'\tilde{\Sigma}_t^{-1}s \leq \max_s s'(\Sigma_t^*)^{-1}s \leq \max\big(\max_s s'\Sigma_a^{-1}s, \max_s s'\Sigma_b^{-1}s\big) \tag{A.18}$$

(where $\max_s$ is the maximum over $s = (s_1, \ldots, s_k) \in \{-1, 1\}^k$), so that both the mappings $\Sigma \mapsto \max(\mathrm{diag}(\Sigma))$ and $\Sigma \mapsto \max_s s'\Sigma^{-1}s$ are geodesic and harmonic quasi-convex.

Proof of Lemma A.9. The result in (A.17) readily follows from the fact that the weighted harmonic-geometric-arithmetic inequality $\Sigma_t^* \leq \tilde{\Sigma}_t \leq (1-t)\Sigma_a + t\Sigma_b$ yields $(\Sigma_t^*)_{\ell\ell} \leq (\tilde{\Sigma}_t)_{\ell\ell} \leq (1-t)(\Sigma_a)_{\ell\ell} + t(\Sigma_b)_{\ell\ell} \leq \max((\Sigma_a)_{\ell\ell}, (\Sigma_b)_{\ell\ell})$ for any $\ell = 1, \ldots, k$. Turning to (A.18), the harmonic-geometric inequality implies that $\tilde{\Sigma}_t^{-1} \leq (\Sigma_t^*)^{-1}$, which readily yields $\max_s s'\tilde{\Sigma}_t^{-1}s \leq \max_s s'(\Sigma_t^*)^{-1}s$. Consequently, it only remains to prove the second inequality in (A.18). To do so, choose an arbitrary s_* such that $\max_s s'(\Sigma_t^*)^{-1}s = s_*'(\Sigma_t^*)^{-1}s_*$. Then

$$\begin{aligned}
\max_s s'(\Sigma_t^*)^{-1}s = s_*'(\Sigma_t^*)^{-1}s_* &= s_*'\big((1-t)\Sigma_a^{-1} + t\Sigma_b^{-1}\big)s_* \\
&\leq (1-t)\max_s s'\Sigma_a^{-1}s + t\max_s s'\Sigma_b^{-1}s \\
&\leq \max\big(\max_s s'\Sigma_a^{-1}s, \max_s s'\Sigma_b^{-1}s\big),
\end{aligned} \tag{A.19}$$

which establishes the result. □

We can now prove Theorems 5.2 and 6.1.

Proof of Theorem 5.2. (i) We start by considering the case where P is an elliptical probability measure over $\mathbb{R}^k$ with location θ_0 and scatter Σ_0, where we first prove the result for $\theta_0 = 0$ and $\Sigma_0 = I_k$. Then we have

$$\begin{aligned}
HD_{P,T}^{\mathrm{sc}}(\tilde{\Sigma}_t) &= \inf_{u \in \mathcal{S}^{k-1}} \min\Big(P\big[|u'X| \leq \sqrt{u'\tilde{\Sigma}_t u}\,\big], P\big[|u'X| \geq \sqrt{u'\tilde{\Sigma}_t u}\,\big]\Big) \\
&= \min\Big(\inf_{u \in \mathcal{S}^{k-1}} P[|X_1| \leq \sqrt{u'\tilde{\Sigma}_t u}], \inf_{u \in \mathcal{S}^{k-1}} P[|X_1| \geq \sqrt{u'\tilde{\Sigma}_t u}]\Big) \\
&= \min(P[|X_1| \leq \lambda_k^{1/2}(\tilde{\Sigma}_t)], P[|X_1| \geq \lambda_1^{1/2}(\tilde{\Sigma}_t)]).
\end{aligned} \tag{A.20}$$

Since Lemma A.8 entails that

$$\begin{aligned}
P[|X_1| \leq \lambda_k^{1/2}(\tilde{\Sigma}_t)] &\geq P[|X_1| \leq \min(\lambda_k^{1/2}(\Sigma_a), \lambda_k^{1/2}(\Sigma_b))] \\
&= \min(P[|X_1| \leq \lambda_k^{1/2}(\Sigma_a)], P[|X_1| \leq \lambda_k^{1/2}(\Sigma_b)]) \\
&\geq \min(HD_{P,T}^{\mathrm{sc}}(\Sigma_a), HD_{P,T}^{\mathrm{sc}}(\Sigma_b))
\end{aligned}$$

and

$$\begin{aligned}
P[|X_1| \geq \lambda_1^{1/2}(\tilde{\Sigma}_t)] &\geq P[|X_1| \geq \max(\lambda_1^{1/2}(\Sigma_a), \lambda_1^{1/2}(\Sigma_b))] \\
&= \min(P[|X_1| \geq \lambda_1^{1/2}(\Sigma_a)], P[|X_1| \geq \lambda_1^{1/2}(\Sigma_b)]) \\
&\geq \min(HD_{P,T}^{\mathrm{sc}}(\Sigma_a), HD_{P,T}^{\mathrm{sc}}(\Sigma_b)),
\end{aligned}$$

the result for $\theta_0 = 0$ and $\Sigma_0 = I_k$ follows from (A.20).

We now prove the result in the elliptical case with arbitrary values of θ_0 and Σ_0. To this end, let $A = \Sigma_0^{-1/2}$ and note that the square roots (in $\mathcal{P}_k$) of $\Upsilon_a := A\Sigma_a A'$ and $\Upsilon_b := A\Sigma_b A'$ are of the form $\Upsilon_a^{1/2} = A\Sigma_a^{1/2} O_a$ and $\Upsilon_b^{1/2} = A\Sigma_b^{1/2} O_b$, for some $k \times k$ orthogonal matrices O_a, O_b. Consequently,

$$\begin{aligned}\tilde{\Upsilon}_t &:= A\tilde{\Sigma}_t A' = A\Sigma_a^{1/2}(\Sigma_a^{-1/2}\Sigma_b\Sigma_a^{-1/2})^t\Sigma_a^{1/2}A' \\ &= \Upsilon_a^{1/2}O_a'(O_a\Upsilon_a^{-1/2}\Upsilon_b\Upsilon_a^{-1/2}O_a')^t O_a\Upsilon_a^{1/2} = \Upsilon_a^{1/2}(\Upsilon_a^{-1/2}\Upsilon_b\Upsilon_a^{-1/2})^t\Upsilon_a^{1/2}\end{aligned}$$

describes a geodesic path from Υ_a to Υ_b. Since the result holds at $P_0 = P_{A,-A\theta_0}$ (where the notation $P_{A,b}$ was defined on page 16 of the main manuscript), affine invariance then entails that

$$\begin{aligned}HD^{\mathrm{sc}}_{P,T}(\tilde{\Sigma}_t) = HD^{\mathrm{sc}}_{P_0,T}(\tilde{\Upsilon}_t) &\geq \min(HD^{\mathrm{sc}}_{P_0,T}(\Upsilon_a), HD^{\mathrm{sc}}_{P_0,T}(\Upsilon_b)) \\ &= \min(HD^{\mathrm{sc}}_{P,T}(\Sigma_a), HD^{\mathrm{sc}}_{P,T}(\Sigma_b)),\end{aligned}$$

as was to be showed.

We now turn to the case where the probability measure P over $\mathbb{R}^k$ has independent Cauchy marginals. Fix $\Sigma_a, \Sigma_b \in \mathcal{P}_k$ and consider the geodesic path $\tilde{\Sigma}_t$, $t \in [0,1]$, from Σ_a to Σ_b. Recall that

$$HD^{\mathrm{sc}}_{P,T}(\Sigma) = 2\min\big(\Psi\big(1/\max_s \sqrt{s'\Sigma^{-1}s}\,\big) - \tfrac{1}{2}, 1 - \Psi\big(\sqrt{\max(\mathrm{diag}(\Sigma))}\,\big)\big),$$

where Ψ stands for the Cauchy cumulative distribution function; see (2.7). Lemma A.9 readily entails that

$$\begin{aligned}&2 - 2\Psi\big(\sqrt{\max(\mathrm{diag}(\tilde{\Sigma}_t))}\,\big) \\ &\quad\geq \min\big(2 - 2\Psi\big(\sqrt{\max(\mathrm{diag}(\Sigma_a))}\,\big), 2 - 2\Psi\big(\sqrt{\max(\mathrm{diag}(\Sigma_b))}\,\big)\big) \\ &\quad\geq \min(HD^{\mathrm{sc}}_{P,T}(\Sigma_a), HD^{\mathrm{sc}}_{P,T}(\Sigma_b)).\end{aligned} \tag{A.21}$$

Lemma A.9 also provides $\max_s s'\tilde{\Sigma}_t^{-1}s \leq \max(\max_s s'\Sigma_a^{-1}s, \max_s s'\Sigma_b^{-1}s)$, which rewrites

$$1/\max_s (s'\tilde{\Sigma}_t^{-1}s)^{1/2} \geq \min(1/\max_s (s'\Sigma_a^{-1}s)^{1/2}, 1/\max_s (s'\Sigma_b^{-1}s)^{1/2}).$$

This implies that

$$\begin{aligned}&2\Psi\big(1/\max_s (s'\tilde{\Sigma}_t^{-1}s)^{1/2}\,\big) - 1 \\ &\quad\geq \min\big(2\Psi\big(1/\max_s (s'\Sigma_a^{-1}s)^{1/2}\,\big) - 1, 2\Psi\big(1/\max_s (s'\Sigma_b^{-1}s)^{1/2}\,\big) - 1\big) \\ &\quad\geq \min(HD^{\mathrm{sc}}_{P,T}(\Sigma_a), HD^{\mathrm{sc}}_{P,T}(\Sigma_b)).\end{aligned} \tag{A.22}$$

From (A.21)-(A.22), it readily follows that $HD^{\mathrm{sc}}_{P,T}(\tilde{\Sigma}_t) \geq \min(HD^{\mathrm{sc}}_{P,T}(\Sigma_a), HD^{\mathrm{sc}}_{P,T}(\Sigma_b))$, which concludes the proof of Part (i).

(ii) Let then P be an arbitrary probability measure over $\mathbb{R}^k$ satisfying Part (i) of the result.

For any $\Sigma_a, \Sigma_b \in R^{\rm sc}_{P,T}(\alpha)$, we then have $HD^{\rm sc}_{P,T}(\tilde{\Sigma}_t) \geq \min(HD^{\rm sc}_{P,T}(\Sigma_a), HD^{\rm sc}_{P,T}(\Sigma_b)) \geq \alpha$, so that $\tilde{\Sigma}_t \in R^{\rm sc}_{P,T}(\alpha)$. $\square$

Proof of Theorem 6.1. In view of the remark given right before the statement of the theorem, it is sufficient to show that, at the probability measures P considered, $\Sigma \mapsto HD^{\rm sc}_{P,T}(\Sigma)$ is harmonic quasi-concave. The proof is then entirely similar to the proof of Theorem 5.2. In the elliptical case, the affine-invariance argument is based on the identity $\Upsilon^*_t := A\Sigma^*_t A' = ((1-t)\Upsilon_a^{-1} + t\Upsilon_b^{-1})^{-1}$, with $\Upsilon_a := A\Sigma_a A'$ and $\Upsilon_b := A\Sigma_b A'$. $\square$

A.5 Proofs from Section 7

Proof of Theorem 7.1. We may restrict to the case where $HD^{{\rm sh},S}_{P,T}(V) > \alpha_{P,T}$ (indeed, the assumptions ensure that $HD^{{\rm sh},S}_{P,T}(V) \geq \alpha_{P,T}$ and that the result holds if $HD^{{\rm sh},S}_{P,T}(V) = \alpha_{P,T}$). For any $\alpha > \alpha_{P,T}$, consider then $I_\alpha := I_{\alpha,P,T}(V) := \{\sigma^2 \in \mathbb{R}_0^+ : \sigma^2 V \in R^{\rm sc}_{P,T}(\alpha)\} = \{\sigma^2 \in \mathbb{R}_0^+ : HD^{\rm sc}_{P,T}(\sigma^2 V) \geq \alpha\}$. The convexity of $R^{\rm sc}_{P,T}(\alpha)$ (Theorem 3.3(ii)) implies that I_α is an interval. Since $\alpha > \alpha_{P,T}$, Theorem 4.2 shows that $R^{\rm sc}_{P,T}(\alpha)$ is g-bounded, which implies there exist $\eta_\alpha > 0$ and $M_\alpha > \eta_\alpha$ such that $I_\alpha \subset [\eta_\alpha, M_\alpha]$. Since, moreover, Theorem 3.1 implies that $\sigma^2 \mapsto HD^{\rm sc}_{P,T}(\sigma^2 V)$ is upper semicontinuous, I_α is also closed, hence (still for $\alpha > \alpha_{P,T}$) compact.

Now, fix $\delta > 0$ such that $HD^{{\rm sh},S}_{P,T}(V) - \delta > \alpha_{P,T}$. For any n, pick then σ_n^2 in the (non-empty) interval $I_{HD^{{\rm sh},S}_{P,T}(V)-(\delta/n)}$. The resulting sequence (σ_n^2) is in the compact set $I_{HD^{{\rm sh},S}_{P,T}(V)-\delta}$, hence admits a subsequence $(\sigma^2_{n_\ell})$ converging in $\mathbb{R}_0^+$, to σ_V^2, say. Fix then an arbitrary $\varepsilon \in (0,\delta)$. For ℓ large enough, all $\sigma^2_{n_\ell}$ belong to the closed set $I_{HD^{{\rm sh},S}_{P,T}(V)-\varepsilon}$, so that σ_V^2 also belongs to $I_{HD^{{\rm sh},S}_{P,T}(V)-\varepsilon}$. This shows that $HD^{{\rm sh},S}_{P,T}(V) - \varepsilon \leq HD^{\rm sc}_{P,T}(\sigma_V^2 V) \leq HD^{{\rm sh},S}_{P,T}(V)$. Since ε can be taken arbitrarily small, the result is proved. $\square$

Proof of Theorem 7.2. From Theorem 2.1, we readily obtain

$$
\begin{aligned}
HD^{{\rm sh},S}_{P_{A,b},T}(AVA'/S(AVA')) &= \sup_{\sigma^2>0} HD^{\rm sc}_{P_{A,b},T}(\sigma^2 AVA'/S(AVA')) \\
&= \sup_{\sigma^2>0} HD^{\rm sc}_{P_{A,b},T}(\sigma^2 AVA') = \sup_{\sigma^2>0} HD^{\rm sc}_{P,T}(\sigma^2 V) = HD^{{\rm sh},S}_{P,T}(V),
\end{aligned}
$$

which establishes the result. $\square$

Proof of Theorem 7.3. Consider arbitrary probability measures P, Q on $\mathbb{R}^k$. Fix $V \in \mathcal{P}_k^S$ and assume (without loss of generality) that $HD^{{\rm sh},S}_{P,T}(V) \leq HD^{{\rm sh},S}_{Q,T}(V)$. Then, for any $\varepsilon > 0$, there exists $\sigma_\varepsilon^2 > 0$ such that $HD^{{\rm sh},S}_{Q,T}(V) \leq HD^{\rm sc}_{Q,T}(\sigma_\varepsilon^2 V) + \varepsilon$, so that

$$
\begin{aligned}
|HD^{{\rm sh},S}_{P,T}(V) - HD^{{\rm sh},S}_{Q,T}(V)| &= HD^{{\rm sh},S}_{Q,T}(V) - HD^{{\rm sh},S}_{P,T}(V) \\
&\leq HD^{\rm sc}_{Q,T}(\sigma_\varepsilon^2 V) + \varepsilon - HD^{\rm sc}_{P,T}(\sigma_\varepsilon^2 V) \leq \sup_{\Sigma \in \mathcal{P}_k} |HD^{\rm sc}_{P,T}(\Sigma) - HD^{\rm sc}_{Q,T}(\Sigma)| + \varepsilon.
\end{aligned}
$$

Since this holds for any $\varepsilon > 0$ and since V is arbitrary, we have that

$$\sup_{V\in\mathcal{P}_k^S} |HD^{\mathrm{sh},S}_{P,T}(V) - HD^{\mathrm{sh},S}_{Q,T}(V)| \leq \sup_{\Sigma\in\mathcal{P}_k} |HD^{\mathrm{sc}}_{Q,T}(\Sigma) - HD^{\mathrm{sc}}_{P,T}(\Sigma)|.$$

The result then follows from Theorem 2.2. □

Proof of Theorem 7.4. (i) Fix a shape matrix $V_0 \in R^{\mathrm{sh},S}_{P,T}(\alpha_{P,T})$ and assume, ad absurdum, that there exists a sequence (V_n) in $\mathcal{P}_k^S$ that g-converges (resp., F-converges) to V_0 and such that

$$\limsup_{n\to\infty} HD^{\mathrm{sh},S}_{P,T}(V_n) > HD^{\mathrm{sh},S}_{P,T}(V_0).$$

Extracting a subsequence if necessary, we can fix $\varepsilon > 0$ small enough to have $HD^{\mathrm{sh},S}_{P,T}(V_0)+\varepsilon < HD^{\mathrm{sh},S}_{P,T}(V_n)$ for any n. Fix then, for any n, $\sigma_n^2 > 0$ such that $HD^{\mathrm{sc}}_{P,T}(\sigma_n^2 V_n) > HD^{\mathrm{sh},S}_{P,T}(V_n)-\varepsilon/2$, which yields

$$HD^{\mathrm{sc}}_{P,T}(\sigma_n^2 V_n) > HD^{\mathrm{sh},S}_{P,T}(V_0) + \varepsilon/2 \geq \alpha_{P,T} + \varepsilon/2. \tag{A.23}$$

Now, we can assume without loss of generality that V_n belongs to a neighbourhood of V_0 that is g-compact in $\mathcal{P}_k^S(P)$. Since (A.23) implies that $\sigma_n^2 V_n$ belongs, for any n, to the g-bounded (Theorem 4.2) scatter depth region $R^{\mathrm{sc}}_{P,T}(\alpha_{P,T}+\varepsilon/2)$, the sequence (σ_n^2) then stays away from 0 and ∞ (that is, the σ_n^2's belong to a common compact set of $\mathbb{R}_0^+$). Consequently, there exists a subsequence $(\sigma_{n_\ell}^2)$ such that $(\sigma_{n_\ell}^2 V_{n_\ell})$ g-converges (resp., F-converges) to $\sigma_0^2 V_0$, say. In view of (A.23), we therefore found $\varepsilon > 0$ such that, for any ℓ,

$$HD^{\mathrm{sc}}_{P,T}(\sigma_{n_\ell}^2 V_{n_\ell}) > HD^{\mathrm{sh},S}_{P,T}(\sigma_0^2 V_0) + \varepsilon/2,$$

where $(\sigma_{n_\ell}^2 V_{n_\ell})$ g-converges (resp., F-converges) to $\sigma_0^2 V_0$, which contradicts the scatter depth upper semicontinuity result in Theorem 4.1 (resp., in Theorem 3.1). (ii) The result follows from the fact that $R^{\mathrm{sh}}_{P,T}(\alpha)$ is the inverse image of $[\alpha, +\infty)$ by the upper F- and g-semicontinuous function $V \mapsto HD^{\mathrm{sh},S}_{P,T}(V)$. (iii) Since the supremum of lower semicontinuous functions is a lower semicontinuous function, Theorems 3.1 and 4.1(iii) yield that $V \mapsto HD^{\mathrm{sh},S}_{P,T}(V)$ is lower-semi continuous. The result then follows from Part (i) and the fact that the smoothness of P at T_P implies that $R^{\mathrm{sh},S}_{P,T}(\alpha_{P,T}) = R^{\mathrm{sh},S}_{P,T}(0) = \mathcal{P}_k^S$. □

The proof of Theorem 7.5 requires the following lemma.

Lemma A.10. Let S be a scale functional, that is a mapping from $\mathcal{P}_k$ to $\mathbb{R}_0^+$ that satisfies the properties (i)-(iii) on page 31 of the main manuscript. Then, $\lambda_k(V) \leq 1 \leq \lambda_1(V)$ for any $V \in \mathcal{P}_k^S$.

Proof of Lemma A.10. (a) Writing the factorisation of V as $V = O\,\mathrm{diag}(\lambda_1(V), \ldots, \lambda_k(V))O'$, where O is a $k \times k$ orthogonal matrix, it holds

$$\lambda_k(V)I_k = O\,\mathrm{diag}(\lambda_k(V), \ldots, \lambda_k(V))O' \leq V \leq O\,\mathrm{diag}(\lambda_1(V), \ldots, \lambda_1(V))O' = \lambda_1(V)I_k$$

(where $A \leq B$ still means that $B - A$ is positive semidefinite), the properties of a scale functional yield $\lambda_k(V) = S(\lambda_k(V)I_k) \leq S(V) \leq S(\lambda_1(V)I_k) = \lambda_1(V)$. □

Proof of Theorem 7.5. We start with the proof of the result for $s_{P,T} < 1/2$ and g-boundedness. We fix $\varepsilon > 0$ and intend to prove that $R^{\mathrm{sh},S}_{P,T}(s_{P,T}+\varepsilon)$ is g-bounded by showing that, for $r > 0$ large enough, it is included in the g-ball $B_g(I_k, r)$. To do so, fix a shape matrix $V (\in \mathcal{P}^S_k)$ that does not belong to $B_g(I_k, r)$ (r is to be chosen later). In view of (A.10), we then have (i) $\lambda_1(V) > \exp(r/\sqrt{k})$ or (ii) $\lambda_k(V) < \exp(-r/\sqrt{k})$ (or both).

We start with case (i). Fix (so far, arbitrarily) $\sigma_0^2 > 0$. Then for any $\sigma^2 \in (0, \sigma_0^2]$, Lemma A.10 entails that (denoting by $v_k(V)$ an arbitrary unit vector associated with $\lambda_k(V)$)

$$\begin{aligned} HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) &\leq \inf_{u\in\mathcal{S}^{k-1}} P\big[|u'(X-T_P)| \leq \sigma\sqrt{u'Vu}\,\big] \\ &\leq P\big[|v_k'(V)(X-T_P)| \leq \sigma\lambda_k^{1/2}(V)\big] \\ &\leq P[|v_k'(V)(X-T_P)| \leq \sigma_0] \leq s_P^{\{T_P\}}(\sigma_0), \end{aligned} \tag{A.24}$$

where we used the notation $s_P^K(\cdot)$ introduced in Lemma A.3. By using this lemma, pick then $\sigma_0^2 > 0$ such that $s_P^{\{T_P\}}(\sigma_0) < s_P^{\{T_P\}} + (\varepsilon/2) = s_{P,T} + (\varepsilon/2)$. Denoting by $v_1(V)$ an arbitrary unit vector associated with $\lambda_1(V)$, we then have that, for any $\sigma^2 \in [\sigma_0^2, \infty)$,

$$\begin{aligned} HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) &\leq \inf_{u\in\mathcal{S}^{k-1}} P\big[|u'(X-T_P)| \geq \sigma\sqrt{u'Vu}\,\big] \\ &\leq P\big[|v_1'(V)(X-T_P)| \geq \sigma\lambda_1^{1/2}(V)\,\big] \\ &\leq P\big[|v_1'(V)(X-T_P)| \geq \sigma_0\exp(r/2\sqrt{k})\big] \\ &\leq P\big[\|X-T_P\| \geq \sigma_0\exp(r/2\sqrt{k})\big], \end{aligned} \tag{A.25}$$

which, for r large enough, can be made smaller than $\varepsilon/2$ (hence, smaller than $s_{P,T} + (\varepsilon/2)$). For r large enough, thus, (A.24)-(A.25) guarantee that $HD^{\mathrm{sh},S}_{P,T}(V) = \sup_{\sigma^2>0} HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) < s_{P,T} + \varepsilon$, as was to be showed.

We then turn to case (ii). By picking σ_0 large enough, we have that, for any $\sigma^2 \in [\sigma_0^2, \infty)$,

$$\begin{aligned} HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) &\leq \inf_{u\in\mathcal{S}^{k-1}} P\big[|u'(X-T_P)| \geq \sigma\sqrt{u'Vu}\,\big] \\ &\leq P\big[|v_1'(V)(X-T_P)| \geq \sigma\lambda_1^{1/2}(V)\,\big] \\ &\leq P\big[|v_1'(V)(X-T_P)| \geq \sigma_0\big] \\ &\leq P\big[\|X-T_P\| \geq \sigma_0\big] < \varepsilon/2, \end{aligned} \tag{A.26}$$

where we used Lemma A.10. For any $\sigma^2 \in (0, \sigma_0^2]$, we then have

$$\begin{aligned} HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) &\leq \inf_{u\in\mathcal{S}^{k-1}} P\big[|u'(X-T_P)| \leq \sigma\sqrt{u'Vu}\,\big] \\ &\leq P\big[|v_k'(V)(X-T_P)| \leq \sigma\lambda_k^{1/2}(V)\big] \\ &\leq P[|v_k'(V)(X-T_P)| \leq \sigma_0\exp(-r/2\sqrt{k})] \\ &\leq s_P^{\{T_P\}}(\sigma_0\exp(-r/2\sqrt{k}) < s_{P,T} + (\varepsilon/2), \end{aligned} \tag{A.27}$$

for r large enough. Thus, for r large enough, (A.26)-(A.27) still yield that $HD^{\mathrm{sh},S}_{P,T}(V) = \sup_{\sigma^2>0} HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) < s_{P,T} + \varepsilon$, as was to be showed. We thus conclude that, for $\alpha > s_{P,T}$, $R^{\mathrm{sh},S}_{P,T}(\alpha)$ is g-bounded (its g-compacity then follows from the same argument as in the proof of Theorem 4.2).

The proof for F-boundedness (still for $s_P < 1/2$) follows along the same lines and is actually simpler since only one of both cases (i)-(ii) above is to be considered. Recall indeed that, as seen in the proof of Theorem 3.2, $V \notin B_F(I_k, r)$ implies that $\lambda_1(V) > (r-1)/k^{1/2}$, so that the same reasoning as in case (i) above allows to show that for any $\varepsilon > 0$, there exists $r = r(\varepsilon)$ such that $V \notin B(I_k, r)$ implies $HD^{\mathrm{sh},S}_{P,T}(V) < s_{P,T} + \varepsilon$. This establishes that $R^{\mathrm{sh},S}_{P,T}(\alpha)$ is F-bounded for $\alpha > s_{P,T}$.

Finally, if $s_{P,T} \geq 1/2$, then, with $u_0 \in \mathcal{S}^{k-1}$ such that $P[|u_0'(X - T_P)| = 0] = s_{P,T}$ (existence is guaranteed in Lemma A.3(ii), with $K = \{T_P\}$), we have $HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) \leq P[|u_0'(X - T_P)| \geq \sigma\sqrt{u_0' V u_0}\,] \leq P[|u_0'(X - T_P)| > 0] = 1 - s_{P,T}$ for any $\sigma^2 > 0$, so that $HD^{\mathrm{sh},S}_{P,T}(V) \leq 1 - s_{P,T}$. Therefore, $R^{\mathrm{sh},S}_{P,T}(\alpha)$ is empty for any $\alpha > 1 - s_{P,T} = \alpha_{P,T}$. □

Proof of Theorem 7.6. The proof follows along the exact same lines as that of Theorem 4.3, hence is not reported here. □

Proof of Theorem 7.7. (i) For any $V \in \mathcal{P}_k^S$, Theorem 5.1(i) readily implies that $HD^{\mathrm{sh},S}_{P,T}(V) = \sup_{\sigma^2>0} HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) \leq HD^{\mathrm{sc}}_{P,T}(\Sigma_0)$. Since $HD^{\mathrm{sc}}_{P,T}(\Sigma_0) = HD^{\mathrm{sc}}_{P,T}(S(\Sigma_0)V_0) \leq HD^{\mathrm{sh},S}_{P,T}(V_0)$, the result follows. (ii) Before proceeding, note that since P is an elliptical probability measure with location θ_0, the affine-equivariance of T implies that $T_P = \theta_0$. Ellipticity further entails that $s_{P,T} = P[\{\theta_0\}]$. Moreover, we must have $s_{P,T} < 1/2$ (otherwise, $P[|Z_1| = 0] \geq P[\{\theta_0\}] = s_{P,T} \geq 1/2$, so that $0 \in \mathcal{I}_{\mathrm{MSD}}[Z_1]$, a contradiction). Now, assume, ad absurdum, that there exists $V \in \mathcal{P}_k^S \setminus \{V_0\}$ with $HD^{\mathrm{sh},S}_{P,T}(V) = HD^{\mathrm{sh},S}_{P,T}(V_0)$. Since $HD^{\mathrm{sh},S}_{P,T}(V_0) = HD^{\mathrm{sc}}_{P,T}(\Sigma_0) \geq 1/2$, we must have $HD^{\mathrm{sh},S}_{P,T}(V) \geq 1/2$. Since $\alpha_{P,T} = s_{P,T} < 1/2$, there exists $\sigma^2 > 0$ such that $HD^{\mathrm{sc}}_{P,T}(\sigma^2 V) \geq s_P$. Therefore, Theorem 7.1 ensures that $HD^{\mathrm{sh},S}_{P,T}(V) = HD^{\mathrm{sc}}_{P,T}(\sigma_V^2 V)$ for some $\sigma_V^2 > 0$. We therefore have

$$HD^{\mathrm{sc}}_{P,T}(\sigma_V^2 V) = HD^{\mathrm{sh},S}_{P,T}(V) = HD^{\mathrm{sh},S}_{P,T}(V_0) = HD^{\mathrm{sc}}_{P,T}(\Sigma_0) = HD^{\mathrm{sc}}_{P,T}(S(\Sigma_0)V_0).$$

Theorem 5.1(ii) then yields that $\mathrm{Sp}((S(\Sigma_0))^{-1/2}\sigma_V V_0^{-1/2} V^{1/2}) \subset \mathcal{I}_{\mathrm{MSD}}[Z_1]$. Since, by assumption, $\mathcal{I}_{\mathrm{MSD}}[Z_1] = \{1\}$, $V_0^{-1/2} V^{1/2}$ must be proportional to the identity matrix, which implies that $V = V_0$. □

Proof of Theorem 7.8. (i) Consider the scale functional S_{tr} and fix $\varepsilon > 0$. By definition, there exist positive real numbers σ_a^2 and σ_b^2 such that

$$HD^{\mathrm{sc}}_{P,T}(\sigma_a^2 V_a) \geq HD^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(V_a) - \varepsilon \quad \text{and} \quad HD^{\mathrm{sc}}_{P,T}(\sigma_b^2 V_b) \geq HD^{\mathrm{sh},S_{\mathrm{tr}}}_{P,T}(V_b) - \varepsilon.$$

Consider then the linear path $\Sigma_t = (1-t)\Sigma_a + t\Sigma_b$ from $\Sigma_a = \sigma_a^2 V_a$ to $\Sigma_b = \sigma_b^2 V_b$. Letting $h(t) =$

$t\sigma_b^2/((1-t)\sigma_a^2 + t\sigma_b^2)$, the $S_{\rm tr}$-shape matrix associated with Σ_t is

$$\frac{k}{\mathrm{tr}[\Sigma_t]}\,\Sigma_t = \frac{k}{\mathrm{tr}[(1-t)\sigma_a^2 V_a + t\sigma_b^2 V_b]}\,((1-t)\sigma_a^2 V_a + t\sigma_b^2 V_b)$$
$$= \frac{k}{\mathrm{tr}[(1-h(t))V_a + h(t)V_b]}\,((1-h(t))V_a + h(t)V_b) = V_{h(t)}.$$

Since $h : [0,1] \to [0,1]$ is a one-to-one mapping, Theorem 3.3 yields that

$$HD_{P,T}^{\mathrm{sh},S_{\rm tr}}(V_t, P) = HD_{P,T}^{\mathrm{sh},S_{\rm tr}}\Big(\frac{k}{\mathrm{tr}[\Sigma_{h^{-1}(t)}]}\,\Sigma_{h^{-1}(t)}, P\Big) \geq HD_{P,T}^{\mathrm{sc}}(\Sigma_{h^{-1}(t)})$$
$$\geq \min(HD_{P,T}^{\mathrm{sc}}(\Sigma_a), HD_{P,T}^{\mathrm{sc}}(\Sigma_b)) \geq \min(HD_{P,T}^{\mathrm{sh},S_{\rm tr}}(V_a), HD_{P,T}^{\mathrm{sh},S_{\rm tr}}(V_b)) - \varepsilon,$$

for any $t \in [0,1]$. Since this holds for any $\varepsilon > 0$, Part (i) of the result is proved for $S = S_{\rm tr}$. The proof for $S = S_{11}$ is along the exact same lines, hence is omitted. As for Part (ii), it strictly follows like Part (ii) of Theorem 5.2. □

Proof of Theorem 7.9. (i) Fix $\varepsilon > 0$. By definition, there exist $\sigma_a^2 > 0$ and $\sigma_b^2 > 0$ such that

$$HD_{P,T}^{\mathrm{sc}}(\sigma_a^2 V_a) \geq HD_{P,T}^{\mathrm{sh},S_{\rm det}}(V_a) - \varepsilon \quad \text{and} \quad HD_{P,T}^{\mathrm{sc}}(\sigma_b^2 V_b) \geq HD_{P,T}^{\mathrm{sh},S_{\rm det}}(V_b) - \varepsilon.$$

Consider then the geodesic path $\tilde{\Sigma}_t = \Sigma_a^{1/2}\big(\Sigma_a^{-1/2}\Sigma_b\Sigma_a^{-1/2}\big)^t\Sigma_a^{1/2}$ from $\Sigma_a = \sigma_a^2 V_a$ to $\Sigma_b = \sigma_b^2 V_b$. Then, since $\det\tilde{\Sigma}_t = (\det\Sigma_a)^{1-t}(\det\Sigma_b)^t$, it is easy to check that the $S_{\rm det}$-shape matrix associated with $\tilde{\Sigma}_t$ is $(\det\tilde{\Sigma}_t)^{-1/k}\tilde{\Sigma}_t = V_a^{1/2}\big(V_a^{-1/2}V_bV_a^{-1/2}\big)^tV_a^{1/2} =: \tilde{V}_t$. Therefore, using Theorem 5.2, we obtain

$$HD_{P,T}^{\mathrm{sh},S_{\rm det}}(\tilde{V}_t) \geq HD_{P,T}^{\mathrm{sc}}((\det\tilde{\Sigma}_t)^{1/k}\tilde{V}_t) = HD_{P,T}^{\mathrm{sc}}(\tilde{\Sigma}_t)$$
$$\geq \min(HD_{P,T}^{\mathrm{sc}}(\Sigma_a), HD_{P,T}^{\mathrm{sc}}(\Sigma_b)) \geq \min(HD_{P,T}^{\mathrm{sh},S_{\rm det}}(V_a), HD_{P,T}^{\mathrm{sh},S_{\rm det}}(V_b)) - \varepsilon.$$

Part (i) of the result follows since $\varepsilon > 0$ is arbitrary above. As for Part (ii), it is obtained again as in Part (ii) of Theorem 5.2. □

Proof of Theorem 7.10. (i) Fix $\varepsilon > 0$. By definition, there exist positive real numbers σ_a^2 and σ_b^2 such that

$$HD_{P,T}^{\mathrm{sc}}(\sigma_a^2 V_a) \geq HD_{P,T}^{\mathrm{sh},S^*_{\rm tr}}(V_a) - \varepsilon \quad \text{and} \quad HD_{P,T}^{\mathrm{sc}}(\sigma_b^2 V_b) \geq HD_{P,T}^{\mathrm{sh},S^*_{\rm tr}}(V_b) - \varepsilon.$$

Consider then the harmonic path $\Sigma_t^* = ((1-t)\Sigma_a^{-1} + t\Sigma_b^{-1})^{-1}$ from $\Sigma_a = \sigma_a^2 V_a$ to $\Sigma_b = \sigma_b^{-2}V_b$. Then, letting $h(t) = t\sigma_b^{-2}/((1-t)\sigma_a^{-2} + t\sigma_b^{-2})$, the $S^*_{\rm tr}$-shape matrix associated with Σ_t^* is

$$\frac{\mathrm{tr}[(\Sigma_t^*)^{-1}]}{k}\,\Sigma_t^* = ((1-t)\sigma_a^{-2} + t\sigma_b^{-2})\,((1-t)\sigma_a^{-2}V_a^{-1} + t\sigma_b^{-2}V_b^{-1})^{-1}$$
$$= \big((1-h(t))V_a^{-1} + h(t)V_b^{-1}\big)^{-1} =: V^*_{h(t)}.$$

Since $h : [0,1] \to [0,1]$ is a one-to-one mapping, we obtain that, for any $t \in [0,1]$,

$$HD_{P,T}^{\mathrm{sh},S^*_{\mathrm{tr}}}(V_t^*) = HD_{P,T}^{\mathrm{sh},S^*_{\mathrm{tr}}}\left(\frac{\mathrm{tr}[(\Sigma^*_{h^{-1}(t)})^{-1}]}{k}\,\Sigma^*_{h^{-1}(t)}\right) \geq HD_{P,T}^{\mathrm{sc}}(\Sigma^*_{h^{-1}(t)})$$

$$\geq \min(HD_{P,T}^{\mathrm{sc}}(\Sigma_a), HD_{P,T}^{\mathrm{sc}}(\Sigma_b)) \geq \min(HD_{P,T}^{\mathrm{sh},S^*_{\mathrm{tr}}}(V_a), HD_{P,T}^{\mathrm{sh},S^*_{\mathrm{tr}}}(V_b)) - \varepsilon.$$

Since this holds for any $\varepsilon > 0$, Part (i) of the result is proved. Part (ii) strictly follows like Part (ii) of Theorem 5.2. $\square$

B Proofs from Chapter II

Proofs from Chapter II are detailed in this section. Many of the subsequent results require the following lemma.

Lemma B.1. Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Write $C^M_{\theta,V} = \big\{x \in \mathbb{R}^k \setminus \{\theta\} : (u^x_{\theta,V})' M u^x_{\theta,V} \geq \frac{1}{k}\mathrm{tr}(M)\big\}$ and $\tilde{C}^M_{\theta,V} = \big\{x \in \mathbb{R}^k : (u^x_{\theta,V})' M u^x_{\theta,V} \geq \frac{1}{k}\mathrm{tr}(M)\big\}$, where $u^x_{\theta,V}$ is defined as $V^{-1/2}(x-\theta)/\|V^{-1/2}(x-\theta)\|$ if $x \neq \theta$ and as 0 otherwise. Then, for any $V \in \mathcal{P}_{k,\mathrm{tr}}$ and any $r \in \mathbb{R}$,

$$D_\theta(V,P) = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[\tilde{C}^M_{\theta,V}\big] = \inf_{M \in \mathcal{M}_{k,F}^{\mathrm{all}}} P\big[\tilde{C}^M_{\theta,V}\big] = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[C^M_{\theta,V}\big] = \inf_{M \in \mathcal{M}_k^{r}} P\big[C^M_{\theta,V}\big],$$

where $\mathcal{M}_k^{\mathrm{all}}$ (resp., $\mathcal{M}_k^r$) collects the $k \times k$ symmetric matrices with arbitrary trace (resp., with trace r) and where $\mathcal{M}_{k,F}^{\mathrm{all}}$ is the collection of matrices in $\mathcal{M}_k^{\mathrm{all}}$ with Frobenius norm one.

Proof of Lemma B.1. It directly follows from the definition of Tyler shape depth that

$$D_\theta(V,P) = \inf_{v \in \mathbb{R}^{k^2}} P\big[\big\{x \in \mathbb{R}^k : v'\mathrm{vec}\,(u^x_{\theta,V}(u^x_{\theta,V})' - \tfrac{1}{k}I_k) \geq 0\big\}\big].$$

When v runs over $\mathbb{R}^{k^2}$, the matrix M satisfying $v = \mathrm{vec}(M')$ runs over the collection $\mathcal{N}_k$ of $k \times k$ matrices. Since $(u^x_{\theta,V})' M u^x_{\theta,V} = (u^x_{\theta,V})'\{(M+M')/2\}u^x_{\theta,V}$ for any $M \in \mathcal{N}_k$, this yields

$$\begin{aligned} D_\theta(V,P) &= \inf_{M \in \mathcal{N}_k} P\big[\big\{x \in \mathbb{R}^{k^2} : \mathrm{tr}\big[M\{u^x_{\theta,V}(u^x_{\theta,V})' - \tfrac{1}{k}I_k\}\big] \geq 0\big\}\big] \\ &= \inf_{M \in \mathcal{N}_k} P\big[\tilde{C}^M_{\theta,V}\big] = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[\tilde{C}^M_{\theta,V}\big]. \end{aligned} \tag{B.1}$$

Letting $\mathbb{I}[A]$ be equal to one if condition A holds and to zero otherwise, this provides

$$D_\theta(V,P) = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} \Big(P\big[C^M_{\theta,V}\big] + P[\{\theta\}]\mathbb{I}\big[\mathrm{tr}(M) \leq 0\big]\Big) = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[C^M_{\theta,V}\big], \tag{B.2}$$

where we have used the fact that $P\big[C^M_{\theta,V}\big]$ is unchanged when M is replaced with $M + \lambda I_k$ for any $\lambda \in \mathbb{R}$. The same invariance property explains that the infimum over $\mathcal{M}_k^{\mathrm{all}}$ in (B.2) may be replaced with an infimum over $\mathcal{M}_k^r$ for any r. Finally, the result for $\mathcal{M}_{k,F}^{\mathrm{all}}$ follows from (B.1) by noting that $\tilde{C}^{\lambda M}_{\theta,V} = \tilde{C}^M_{\theta,V}$ for any $\lambda > 0$ and that $M = 0$ cannot provide the infimum in (B.1). The proof is complete. $\square$

B.1 Proofs from Section 2

Proof of Theorem 2.1. (i) Fix $M \in \mathcal{M}_k^{\rm all}$ and consider $\tilde{C}^M = \tilde{C}^M_{0,I_k}$, where $\tilde{C}^M_{\theta,V}$ was defined in Lemma B.1. Since $\tilde{C}^M$ is closed, the mapping $P \mapsto P[\tilde{C}^M]$ is upper semicontinuous for weak convergence. Now, Slutzky's lemma entails that, as $d(V, V_0) \to 0$, the measure defined by $B \mapsto P[\theta + V^{1/2}B]$ converges weakly to the one defined by $B \mapsto P[\theta + V_0^{1/2}B]$. Therefore, $V \mapsto P[\theta + V^{1/2}\tilde{C}^M] = P[\tilde{C}^M_{\theta,V}]$ is upper semicontinuous at V_0. From Lemma B.1, we then obtain that

$$V \mapsto D_\theta(V, P) = \inf_{M \in \mathcal{M}_k^{\rm all}} P[\tilde{C}^M_{\theta,V}],$$

is upper semicontinuous (as the infimum of a collection of upper semicontinuous functions). (ii) The result follows from the fact that the depth region $R_\theta(\alpha, P)$ is the inverse image of $[\alpha, \infty)$ by the upper semicontinuous function $V \mapsto D_\theta(V, P)$. (iii) Fix a sequence (V_n) in $\mathcal{P}_{k,\rm tr}$ such that $d(V_n, V_0) \to 0$. In view of Lemma B.1 again, we can, for any n, pick $M_n (\in \mathcal{M}^{\rm all}_{k,F})$ such that $P[\tilde{C}^{M_n}_{\theta,V_n}] \leq D_\theta(V_n, P) + \frac{1}{n}$. Compactness of $\mathcal{M}^{\rm all}_{k,F}$ ensures that we can extract a subsequence (M_{n_ℓ}) of (M_n) that converges to $M_0 (\in \mathcal{M}^{\rm all}_{k,F})$. Writing $\mathbb{I}[B]$ for the indicator function of the set B, the dominated convergence theorem then yields that

$$P[\tilde{C}^{M_{n_\ell}}_{\theta,V_{n_\ell}}] - P[\tilde{C}^{M_0}_{\theta,V_0}] = \int_{\mathbb{R}^k} (\mathbb{I}[\tilde{C}^{M_{n_\ell}}_{\theta,V_{n_\ell}}] - \mathbb{I}[\tilde{C}^{M_0}_{\theta,V_0}])\, dP \to 0$$

as $\ell \to \infty$ (the absolute continuity assumption on P guarantees that $\mathbb{I}[\tilde{C}^{M_{n_\ell}}_{\theta,V_{n_\ell}}] - \mathbb{I}[\tilde{C}^{M_0}_{\theta,V_0}] \to 0$ P-almost everywhere). Consequently,

$$\liminf_{n\to\infty} D_\theta(V_n, P) = \liminf_{n\to\infty} P[\tilde{C}^M_{\theta,V_n}] = \liminf_{\ell\to\infty} P[\tilde{C}^{M_{n_\ell}}_{\theta,V_{n_\ell}}] = P[\tilde{C}^{M_0}_{\theta,V_0}] \geq D_\theta(V_0, P).$$

We conclude that, if P is absolutely continuous with respect to the Lebesgue measure, then $V \to D_\theta(V, P)$ is also lower semicontinuous, hence continuous. □

The proof of Theorem 2.2 requires the following result.

Lemma B.2. Let P be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Write $u^x_\theta = (x-\theta)/\|x-\theta\|$ if $x \neq \theta$ and 0 otherwise. For any $c \geq 0$, further let $t_{\theta,P}(c) = \sup_{v\in\mathcal{S}^{k-1}} P[|v' u^X_\theta| \leq c]$, so that $t_{\theta,P} = t_{\theta,P}(0) = \sup_{v\in\mathcal{S}^{k-1}} P[v'(X-\theta) = 0]$. Then, $t_{\theta,P}(c) \to t_{\theta,P}$ as $c \to 0$.

Proof of Lemma B.2. Since $t_{\theta,P}(c)$ is increasing in c over $[0, \infty)$ and is larger than or equal to $t_{\theta,P}$ for any positive c, we have that $\tilde{t}_{\theta,P} = \lim_{c\to0} t_{\theta,P}(c)$ exists and is such that $\tilde{t}_{\theta,P} \geq t_{\theta,P}$. Now, fix a decreasing sequence (c_n) converging to 0 and consider an arbitrary sequence (v_n) such that

$$P[|v_n' u^X_\theta| \leq c_n] \geq t_{\theta,P}(c_n) - (1/n).$$

Since $\mathcal{S}^{k-1}$ is compact, we can consider a subsequence (v_{n_ℓ}) that converges to $v_0 (\in \mathcal{S}^{k-1})$; without loss of generality, we can of course assume that this subsequence is such that $(v_0' v_{n_\ell})$ is an increasing sequence. Let then $C_\ell = \{v \in \mathcal{S}^{k-1} : v_0' v \geq v_0' v_{n_\ell}\}$. Clearly, C_ℓ is a decreasing sequence of sets with $\cap_\ell C_\ell = \{v_0\}$, so that

$$\lim_{\ell\to\infty} P[u^X_\theta \in \cup_{v\in C_\ell}\{|v'y| \leq c_{n_\ell}\}] = P[u^X_\theta \in \{|v_0' y| \leq 0\}] = P[|v_0' u^X_\theta| = 0].$$

Now, for any ℓ, we have $P[u_\theta^X \in \cup_{v\in C_\ell}\{|v'y| \leq c_{n_\ell}\}] \geq P[|v_{n_\ell}' u_\theta^X| \leq c_{n_\ell}] \geq t_{\theta,P}(c_{n_\ell}) - (1/n_\ell)$, which implies that $t_{\theta,P} \geq P[|v_0' u_\theta^X| = 0] \geq \tilde{t}_{\theta,P}$. □

Proof of Theorem 2.2. Fix $V \in \mathcal{P}_{k,\mathrm{tr}}$ and denote as $\lambda_1(V)$ (resp., $\lambda_k(V)$) the largest (resp., smallest) eigenvalue of V (possible ties are unimportant below). Letting $v_1(V)$ and $v_k(V)$ be arbitrary corresponding unit eigenvectors, Lemma B.1 provides (with $M_V = v_1(V)v_1'(V) \in \mathcal{M}_k^{\mathrm{all}}$)

$$\begin{aligned} D_\theta(V,P) &\leq P\Big[(u_{\theta,V}^X)' M_V u_{\theta,V}^X \geq \tfrac{1}{k}\mathrm{tr}(M_V), X \neq \theta\Big] \\ &= P\Big[k\lambda_1^{-1}(V)\{v_1'(V)(X-\theta)\}^2 \geq \|V^{-1/2}(X-\theta)\|^2, X\neq\theta\Big] \\ &\leq P\Big[\|V^{-1/2}(X-\theta)\|^2 \leq k\|X-\theta\|^2, X\neq\theta\Big] = P\big[\|V^{-1/2}u_{\theta,I_k}^X\|^2 \leq k, u_\theta^X \neq 0\big], \end{aligned}$$

where we used the inequality $\lambda_1(V) \geq 1$ (which follows from the constraint $\mathrm{tr}(V) = k$) and where u_θ^s is defined in Lemma B.2. Therefore,

$$D_\theta(V,P) \leq P\big[\lambda_1(V^{-1})(v_1'(V)u_\theta^X)^2 \leq k\big] \leq t_{\theta,P}\big((k\lambda_k(V))^{1/2}\big). \tag{B.3}$$

Now, ad absurdum, take $\varepsilon > 0$ such that $R_\theta(t_{\theta,P}+\varepsilon, P)$ is unbounded. This implies that there exists a sequence (V_n) in $\mathcal{P}_{k,\mathrm{tr}}$ satisfying $D_\theta(V_n,P) \geq t_{\theta,P}+\varepsilon$ for any n and for which $d(V_n, I_k) \to \infty$. Since $\lambda_1(V_n) < \mathrm{tr}[V_n] = k$, we must have that $\lambda_k(V_n) \to 0$. Lemma B.2 and (B.3) then imply that $D_\theta(V_n,P) < t_{\theta,P}+\varepsilon$ for n large enough, a contradiction. Consequently, $R_\theta(\alpha,P)$ is bounded for any $\alpha > t_{\theta,P}$.

Now, Lemma A.6 above readily implies that a bounded subset of $\mathcal{P}_{k,\mathrm{tr}}$ is also totally bounded, in the sense that, for any $\varepsilon > 0$, it can be covered by finitely many balls of the form $B(V,\varepsilon) = \{\tilde{V} \in \mathcal{P}_{k,\mathrm{tr}} : d(\tilde{V},V) < \varepsilon\}$. Part (i) of the result and Theorem 2.1(ii) thus entail that, for any $\alpha > t_{\theta,P}$, the region $R_\theta(\alpha,P)$ is closed and totally bounded. The result then follows from the completeness of the metric space $(\mathcal{P}_{k,\mathrm{tr}}, d)$. □

Proof of Theorem 2.3. Let $\alpha_* = \sup_{V\in\mathcal{P}_{k,\mathrm{tr}}} D_\theta(V,P)$. By assumption, $R_\theta(t_{\theta,P},P)$ is non-empty. Thus, $\alpha_* \geq t_{\theta,P}$ and the result holds if $\alpha_* = t_{\theta,P}$. We may therefore assume that $\alpha_* > t_{\theta,P}$. For any n, pick then V_n in $R_\theta((\alpha_* - 1/n)_+, P)$, where $u_+ = \max(u,0)$. Fix $\varepsilon \in (0, \alpha_* - t_{\theta,P})$. For n large enough, all terms of the sequence (V_n) belong to the compact set $R_\theta(\alpha_* - \varepsilon, P)$; see Theorem 2.2. Thus, there exists a subsequence (V_{n_k}) that converges in $R_\theta(\alpha_* - \varepsilon, P)$, to V_* say. For any $\varepsilon' \in (0,\varepsilon)$, all (V_{n_k}) eventually belong to the closed set $R_\theta(\alpha_* - \varepsilon', P)$, so that $V_* \in R_\theta(\alpha_* - \varepsilon', P)$. Therefore, $\alpha_* - \varepsilon' \leq D_\theta(V_*,P) \leq \alpha_*$ for any such ε', which establishes the result. □

The proof of Theorem 2.4 requires the following preliminary result.

Lemma B.3. For any $y \in \mathbb{R}^k$ and any $k\times k$ symmetric matrix M, the mapping $V \mapsto \mathrm{tr}(MV)y'V^{-1}y$ is quasi-convex, that is, for any $V_a, V_b \in \mathcal{P}_{k,\mathrm{tr}}$ and any $t\in[0,1]$, $\mathrm{tr}(MV_t)y'V_t^{-1}y \leq \max\{\mathrm{tr}(MV_a)y'V_a^{-1}y, \mathrm{tr}(MV_b)y'V_b^{-1}y\}$, with $V_t = (1-t)V_a + tV_b$.

Proof of Lemma B.3. We treat two cases separately. (i) Assume first that $\mathrm{tr}(MV_a)\mathrm{tr}(MV_b) > 0$. Write

$$\frac{V_t}{\mathrm{tr}(MV_t)} = (1-s_t)\frac{V_a}{\mathrm{tr}(MV_a)} + s_t\frac{V_b}{\mathrm{tr}(MV_b)}, \quad \text{with } s_t = \frac{t\,\mathrm{tr}(MV_b)}{(1-t)\mathrm{tr}(MV_a) + t\,\mathrm{tr}(MV_b)}.$$

Since $s_t \in [0,1]$, the (weighted) harmonic-arithmetic matrix inequality (see, e.g., Lemma 2.1(vii) in Lawson & Lim, 2013) then shows that, for any $y \in \mathbb{R}^k$,

$$\begin{aligned} y'\left\{\frac{V_t}{\mathrm{tr}(MV_t)}\right\}^{-1} y &\leq y'\left[(1-s_t)\left\{\frac{V_a}{\mathrm{tr}(MV_a)}\right\}^{-1} + s_t\left\{\frac{V_b}{\mathrm{tr}(MV_b)}\right\}^{-1}\right]y \\ &\leq \max\left[y'\left\{\frac{V_a}{\mathrm{tr}(MV_a)}\right\}^{-1} y, y'\left\{\frac{V_b}{\mathrm{tr}(MV_b)}\right\}^{-1} y\right], \end{aligned}$$

as was to be showed. (ii) Assume then that $\mathrm{tr}(MV_a)\mathrm{tr}(MV_b) \leq 0$. Without loss of generality, assume that $\mathrm{tr}(MV_a) \leq 0$ and $\mathrm{tr}(MV_b) \geq 0$. If $\mathrm{tr}(MV_a) = \mathrm{tr}(MV_b) = 0$, then $\mathrm{tr}(MV_t) = 0$ for any t and the result trivially holds. Hence, we may assume that $\mathrm{tr}(MV_a) \neq 0$ or $\mathrm{tr}(MV_b) \neq 0$, which implies that $\mathrm{tr}[MV_{t_0}] = 0$ for a unique $t_0 \in [0,1]$. From continuity, pick then $\delta \in (0, 1-t_0)$ such that, for any $t \in [t_0, t_0+\delta)$,

$$\begin{aligned} \mathrm{tr}(MV_t)y'V_t^{-1}y &\leq \mathrm{tr}(MV_b)y'V_b^{-1}y \\ &\leq \max\{\mathrm{tr}(MV_a)y'V_a^{-1}y, \mathrm{tr}(MV_b)y'V_b^{-1}y\}. \end{aligned}$$

By applying Part (i) of the proof with $V_{t_0+\delta}$ and V_b, we obtain that, for any $t \in [t_0+\delta, 1]$,

$$\begin{aligned} \mathrm{tr}(MV_t)y'V_t^{-1}y &\leq \max\{\mathrm{tr}(MV_{t_0+\delta})y'V_{t_0+\delta}^{-1}y, \mathrm{tr}(MV_b)y'V_b^{-1}y\} \\ &\leq \max\{\mathrm{tr}(MV_a)y'V_a^{-1}y, \mathrm{tr}(MV_b)y'V_b^{-1}y\}. \end{aligned}$$

Since $\mathrm{tr}(MV_t)y'V_t^{-1}y \leq 0 \leq \max\{\mathrm{tr}(MV_a)y'V_a^{-1}y, \mathrm{tr}(MV_b)y'V_b^{-1}y\}$ for any $t \in [0, t_0]$, the result follows. □

Proof of Theorem 2.4. (i) Write $V_t = (1-t)V_a + tV_b$, where $V_a, V_b \in \mathcal{P}_{k,\mathrm{tr}}$ and $t \in [0,1]$ are fixed. First note that, letting $d_\theta^2(V) = (X-\theta)'V^{-1}(X-\theta)$, Lemma B.1 yields

$$\begin{aligned} D_\theta(V,P) &= \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[(X-\theta)'V^{-1/2}MV^{-1/2}(X-\theta) \geq \tfrac{1}{k}\mathrm{tr}(M)d_\theta^2(V), X \neq \theta\big] \\ &= \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[(X-\theta)'M(X-\theta) \geq \tfrac{1}{k}\mathrm{tr}(MV)d_\theta^2(V), X \neq \theta\big]. \end{aligned} \tag{B.4}$$

Writing again $V_t = (1-t)V_a + tV_b$, Lemma B.3 thus yields that, for any $M \in \mathcal{M}_k^{\text{all}}$,

$$\begin{aligned}
P\big[(X-\theta)'M(X-\theta) \geq \tfrac{1}{k}\text{tr}(MV_t)d_\theta^2(V_t), X \neq \theta\big] \\
&\geq P\big[(X-\theta)'M(X-\theta) \geq \tfrac{1}{k}\max\{\text{tr}(MV_a)d_\theta^2(V_a), \text{tr}(MV_b)d_\theta^2(V_b)\}, X \neq \theta\big] \\
&= \min\Big(P\big[(X-\theta)'M(X-\theta) \geq \tfrac{1}{k}\text{tr}(MV_a)d_\theta^2(V_a), X \neq \theta\big], \\
&\qquad\qquad P\big[(X-\theta)'M(X-\theta) \geq \tfrac{1}{k}\text{tr}(MV_b)d_\theta^2(V_b), X \neq \theta\big]\Big) \\
&\geq \min(D_\theta(V_a, P), D_\theta(V_b, P)).
\end{aligned}$$

The result then follows from (B.4). (ii) If $V_a, V_b \in R_\theta(\alpha, P)$, then Part (i) of the result entails that $D_\theta((1-t)V_a + tV_b, P) \geq \min\{D_\theta(V_a, P), D_\theta(V_b, P)\} \geq \alpha$, so that $(1-t)V_a + tV_b \in R_\theta(\alpha, P)$. □

The proof of Theorem 2.5 requires both following lemmas.

Lemma B.4. Let P be elliptical over $\mathbb{R}^k$ with location 0 and shape I_k. Then, $D_0(I_k, P) = (1-P[\{0\}])P[U_1^2 > 1/k]$, where $U = (U_1, \ldots, U_k)'$ is uniformly distributed over the unit sphere $\mathcal{S}^{k-1}$.

Lemma B.5. Let P be elliptical over $\mathbb{R}^k$ with location 0 and shape I_k. Then, for any $V \in \mathcal{P}_{k,\text{tr}} \setminus \{I_k\}$, $D_0(V, P) < (1-P[\{0\}])P[U_1^2 > 1/k]$, where $U = (U_1, \ldots, U_k)'$ is uniformly distributed over $\mathcal{S}^{k-1}$.

Proof of Lemma B.4. In the spherical setup considered, we have that, for any $M \in \mathcal{M}_k^{\text{all}}$,

$$P\bigg[\frac{X'MX}{\|X\|^2} \geq \frac{1}{k}\,\text{tr}(M),\, X \neq 0\bigg] = P\big[U'MU \geq \tfrac{1}{k}\text{tr}(M)\big]P[X \neq 0],$$

where $U = (U_1, \ldots, U_k)'$ is uniform over $\mathcal{S}^{k-1}$. Lemma B.1 then entails that

$$D_0(I_k, P) = (1-P[\{0\}]) \inf_{M \in \mathcal{M}_k^{\text{all}}} P\big[U'MU \geq \tfrac{1}{k}\text{tr}(M)\big].$$

Decomposing M into $O\Lambda O'$, where O is a $k\times k$ orthogonal matrix and where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k)$ is a diagonal matrix, this yields

$$\begin{aligned}
D_0(I_k, P) &= (1-P[\{0\}]) \inf_{\lambda \in \mathbb{R}^k} P\bigg[\sum_{\ell=1}^k \lambda_\ell U_\ell^2 \geq \frac{1}{k}\sum_{\ell=1}^k \lambda_\ell\bigg] \\
&= (1-P[\{0\}]) \inf_{\lambda \in \mathbb{R}^k} P\bigg[\sum_{\ell=1}^k \lambda_\ell\Big(U_\ell^2 - \frac{1}{k}\Big) \geq 0\bigg] =: (1-P[\{0\}]) \inf_{\lambda \in \mathbb{R}^k} p(\lambda).
\end{aligned}$$

By using successively the facts that $p(0) = 1$ and $p(\lambda) = p(\lambda/\|\lambda\|)$ for any $\lambda \in \mathbb{R}^k \setminus \{0\}$, we obtain

$$D_0(I_k, P) = (1-P[\{0\}]) \inf_{\lambda \in \mathbb{R}^k \setminus \{0\}} p(\lambda) = (1-P[\{0\}]) \inf_{\lambda \in \mathcal{S}^{k-1}} p(\lambda). \tag{B.5}$$

The result then follows from Theorem 2 from Paindaveine & Van Bever (2017b), that states that the last infimum in (B.5) is equal to $P[U_1^2 > 1/k]$. □

Proof of Lemma B.5. Fix $V \in \mathcal{P}_{k,\mathrm{tr}}$ and let X be a random k-vector with $P = P^X$. Write $V = O\Lambda O'$, where O is a $k \times k$ orthogonal matrix and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$ is a diagonal matrix with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$. Affine invariance (see Theorem 2.6) entails that

$$D_0(V, P^X) = D_0\big(O'VO, P^{O'X}\big) = D_0\big(\Lambda, P^X\big).$$

Denoting by e_1 the first vector of the canonical basis of $\mathbb{R}^{k^2}$, we then have

$$\begin{aligned} D_0(V, P^X) &= D_0\big(\Lambda, P^X\big) \leq P\big[e_1'\mathrm{vec}\big(U_{0,\Lambda}U_{0,\Lambda}' - \tfrac{1}{k}I_k\big) \geq 0\big] = P\big[((U_{0,\Lambda})_1)^2 - \tfrac{1}{k} \geq 0\big] \\ &= P\big[((U_{0,\Lambda})_1)^2 - \tfrac{1}{k} \geq 0, X \neq 0\big] = P\big[\lambda_1^{-1}X_1^2 \geq \tfrac{1}{k}\textstyle\sum_{\ell=1}^k \lambda_\ell^{-1}X_\ell^2, X \neq 0\big] \\ &\leq P\big[X_1^2 \geq \tfrac{1}{k}\textstyle\sum_{\ell=1}^k X_\ell^2, X \neq 0\big] = P\big[X_1^2/\|X\|^2 \geq \tfrac{1}{k}, X \neq 0\big] = P[X \neq 0]P\big[U_1^2 \geq \tfrac{1}{k}\big], \end{aligned} \tag{B.6}$$

where $U = (U_1, \ldots, U_k)'$ is uniform over $\mathcal{S}^{k-1}$. To have $D_0(V, P^X) = P[X \neq 0]P\big[U_1^2 \geq \frac{1}{k}\big]$, the inequality in (B.6) needs to be an equality, which requires that $\lambda_\ell = \lambda_1$ for all ℓ, hence that $V = I_k$. □

It follows from Rousseeuw & Struyf (2004) that if P is elliptical with location θ, then $M_P = \{\theta\}$, so that, under ellipticity, the Tukey median is Fisher consistent for θ (i.e., $\theta_P = \theta$). We can now prove Theorem 2.5.

Proof of Theorem 2.5. Lemmas B.4-B.5 establish the result in the spherical case associated with $\theta_0 = 0$ and $V_0 = I_k$. For general values of θ_0 and V_0, note that $Y = V_0^{-1/2}(X - \theta_0)$ is elliptical with location 0, shape I_k, and satisfies $P[Y = 0] = P[X = \theta_0]$. By affine invariance,

$$\begin{aligned} D_{\theta_0}(V, P^X) &= D_0\left(\frac{kV_0^{-1/2}VV_0^{-1/2}}{\mathrm{tr}(V_0^{-1/2}VV_0^{-1/2})}, P^Y\right) \\ &\leq D_0(I_k, P^Y) = D_0\left(\frac{kV_0^{-1/2}V_0V_0^{-1/2}}{\mathrm{tr}(V_0^{-1/2}V_0V_0^{-1/2})}, P^Y\right) = D_{\theta_0}(V_0, P^X), \end{aligned}$$

with equality if and only if $kV_0^{-1/2}VV_0^{-1/2}/\mathrm{tr}(V_0^{-1/2}VV_0^{-1/2}) = I_k$, that is, if and only if $V = V_0$. □

Proof of Theorem 2.6. In the proof of Theorem 2.4, we showed that

$$D_\theta(V, P) = \inf_{M \in \mathcal{M}_k^{\mathrm{all}}} P\big[(X - \theta)'V^{-1/2}MV^{-1/2}(X - \theta) \geq \tfrac{1}{k}\mathrm{tr}(M)(X - \theta)'V^{-1}(X - \theta), X \neq \theta\big].$$

Using the fact that $V_A^{1/2} = k^{1/2}AV^{1/2}O/\{\mathrm{tr}(AVA')\}^{1/2}$ for some $k \times k$ orthogonal matrix O (recall that $V_A^{1/2}$ stands for the symmetric positive definite square root of V_A), this readily yields

$$D_{A\theta+b}(V_A, P) = \inf_{M\in\mathcal{M}_k^{\rm all}} P\big[(X-\theta)'V^{-1/2}OMO'V^{-1/2}(X-\theta) \geq \tfrac{1}{k}\mathrm{tr}(OMO')(X-\theta)'V^{-1}(X-\theta), X\neq\theta\big] = D_\theta(V,P),$$

as was to be showed. The affine-equivariance property of the depth regions readily follows. □

The proof of Theorem 2.7 requires the following lemma.

Lemma B.6. For any v_1, v_2 such that $v_1^2+v_2^2<1$, we have

$$\frac{(1-v_1^2)^{1/2}-|v_2|}{(1-v_1^2)^{1/2}+|v_2|} \leq \frac{(1-v_1^2)^{1/2}+|v_2|}{(1-v_1^2)^{1/2}-|v_2|} \leq \frac{1+(v_1^2+v_2^2)^{1/2}}{1-(v_1^2+v_2^2)^{1/2}}.$$

Proof of Lemma B.6. The first inequality is straightforward. For v_1, v_2 such that $v_1^2+v_2^2<1$, let V be the positive definite matrix defined as

$$V = \begin{pmatrix} 1+v_1 & v_2 \\ v_2 & 1-v_1 \end{pmatrix}.$$

Since

$$\frac{\sqrt{1-v_1^2}+v_2}{\sqrt{1-v_1^2}-v_2} = (\det V)^{-1}(\sqrt{1-v_1^2}+v_2)^2 = (\det V)^{-1}\big(1-v_1^2+v_2^2+2v_2\sqrt{1-v_1^2}\big)$$

and

$$\frac{1+\sqrt{v_1^2+v_2^2}}{1-\sqrt{v_1^2+v_2^2}} = (\det V)^{-1}(1+\sqrt{v_1^2+v_2^2})^2 = (\det V)^{-1}(1+v_1^2+v_2^2+2\sqrt{v_1^2+v_2^2}),$$

the second inequality simplifies to $v_2\sqrt{1-v_1^2} \leq v_1^2+\sqrt{v_1^2+v_2^2}$. If $v_2\leq 0$, the proof is complete, so let us focus on the case $v_2>0$. After taking squares, the simplified inequality further yields

$$\begin{aligned} v_2^2(1-v_1^2) \leq v_1^4+(v_1^2+v_2^2)+2v_1^2\sqrt{v_1^2+v_2^2} \;&\Leftrightarrow\; v_2^2-v_1^2v_2^2 \leq v_1^4+v_1^2+v_2^2+2v_1^2\sqrt{v_1^2+v_2^2} \\ &\Leftrightarrow\; -v_1^2v_2^2 \leq v_1^4+v_1^2+2v_1^2\sqrt{v_1^2+v_2^2} \\ &\Leftrightarrow\; 0 \leq v_1^4+v_1^2v_2^2+v_1^2+2v_1^2\sqrt{v_1^2+v_2^2} \\ &\Leftrightarrow\; 0 \leq v_1^2(v_1^2+1)^2+2v_1^2\sqrt{v_1^2+v_2^2}, \end{aligned}$$

which concludes the proof. □

Proof of Theorem 2.7. (i) If $P=P^X$ is elliptical with location θ_0 and shape V_0, then ${V_0}^{-1/2}(X-\theta_0)$ is equal in distribution to RU, where U is uniformly distributed over the unit sphere $\mathcal{S}^{k-1}$

and is independent of the nonnegative random variable R. Theorem 2.6 then yields

$$D_{\theta_0}(V, P^X) = D_0\left(\frac{kV_0^{-1/2}VV_0^{-1/2}}{\mathrm{tr}(V_0^{-1/2}VV_0^{-1/2})}, P^{RU}\right). \tag{B.7}$$

Now, for any $\tilde{V} \in \mathcal{P}_{k,\mathrm{tr}}$, Lemma B.1 entails that

$$\begin{aligned} D_0(\tilde{V}, P^{RU}) &= \inf_{M \in \mathcal{M}_k^0} P\big[U'\tilde{V}^{-1/2}M\tilde{V}^{-1/2}U \geq 0, R > 0\big] \qquad \text{(B.8)} \\ &= P[R > 0] \inf_{M \in \mathcal{M}_k^0} P\big[U'\tilde{V}^{-1/2}M\tilde{V}^{-1/2}U \geq 0\big] = P[R > 0] D_0(\tilde{V}, P^U). \end{aligned}$$

Combining with (B.7), we obtain

$$D_{\theta_0}(V, P^X) = (1 - P^X[\{\theta_0\}]) D_0\left(\frac{kV_0^{-1/2}VV_0^{-1/2}}{\mathrm{tr}(V_0^{-1/2}VV_0^{-1/2})}, P^U\right),$$

which establishes Part (i) of the result. (ii) Assume that $P = P^X$ is bivariate standard normal and fix $V \in \mathcal{P}_{2,\mathrm{tr}}$. We aim at evaluating

$$D_0(V, P^X) = \inf_{M \in \mathcal{M}_k^0} P\big[X'V^{-1/2}MV^{-1/2}X \geq 0\big]; \tag{B.9}$$

see (B.8). To do so, it will be convenient to parametrise V as in Lemma B.6 and the matrix M from as

$$V = \begin{pmatrix} 1 + v_1 & v_2 \\ v_2 & 1 - v_1 \end{pmatrix} \quad \text{and} \quad M = m_1 \begin{pmatrix} 1 & m_2 \\ m_2 & -1 \end{pmatrix},$$

with $v_1^2 + v_2^2 < 1$ and $m_1 \neq 0$ (note indeed that $m_1 = 0$ makes the probability in (B.9) equal to one, which cannot be the infimum). Decomposing $V^{-1/2}MV^{-1/2}$ into $O\Lambda O'$, where O is a 2×2 orthogonal matrix and where $\Lambda = \mathrm{diag}(\lambda_1(V^{-1}M), \lambda_2(V^{-1}M))$, with $\lambda_1(V^{-1}M) \geq \lambda_2(V^{-1}M)$, involves the eigenvalues of $V^{-1}M$ (equivalently, of $V^{-1/2}MV^{-1/2}$), we have

$$D_0(V, P) = \inf_{(m_1, m_2) \in \mathbb{R}_0 \times \mathbb{R}} P\big[\lambda_1(V^{-1}M)X_1^2 + \lambda_2(V^{-1}M)X_2^2 \geq 0\big], \tag{B.10}$$

where $X = (X_1, X_2)'$ is still bivariate standard normal. Since $\lambda_1(-V^{-1}M) = -\lambda_2(V^{-1}M)$ for any $M \in \mathcal{M}_k^0$ (we will show below that $\lambda_2(V^{-1}M) < 0 < \lambda_1(V^{-1}M)$ for any $M \in \mathcal{M}_k^0$), we have

$$\begin{aligned} D_0(V, P) = \min\Big(&\inf_{(m_1, m_2) \in \mathbb{R}_0^+ \times \mathbb{R}} P\big[\lambda_1(V^{-1}M)X_1^2 + \lambda_2(V^{-1}M)X_2^2 \geq 0\big], \qquad \text{(B.11)} \\ &\inf_{(m_1, m_2) \in \mathbb{R}_0^+ \times \mathbb{R}} P\big[\lambda_1(V^{-1}M)X_1^2 + \lambda_2(V^{-1}M)X_2^2 \leq 0\big]\Big), \end{aligned}$$

which allows us to restrict to positive values of m_1. A direct computation shows that, for $m_1 > 0$,

$$\lambda_1(V^{-1}M) = \frac{m_1}{\det V}\left[-(v_1 + m_2 v_2) + \{(v_1 + m_2 v_2)^2 + (1 + m_2^2)\det V\}^{1/2}\right] > 0$$

and

$$\lambda_2(V^{-1}M) = \frac{m_1}{\det V}\left[-(v_1+m_2v_2) - \{(v_1+m_2v_2)^2 + (1+m_2^2)\det V\}^{1/2}\right] < 0.$$

Since $f(m_2) = -\lambda_2(V^{-1}M))/\lambda_1(V^{-1}M)$ does not depend on m_1, (B.11) leads to

$$\begin{aligned} D_0(V,P) &= \min\left(P\left[\frac{X_1^2}{X_2^2} \geq \sup_{m_2\in\mathbb{R}} f(m_2)\right], P\left[\frac{X_1^2}{X_2^2} \leq \inf_{m_2\in\mathbb{R}} f(m_2)\right]\right) \\ &= P\left[\frac{X_1^2}{X_2^2} \geq \max\left(\sup_{m_2\in\mathbb{R}} f(m_2)\ ,\ 1/\inf_{m_2\in\mathbb{R}} f(m_2)\right)\right]. \end{aligned} \tag{B.12}$$

It is easy to check that f is differentiable over $\mathbb{R}$ with a derivative of the form $c_{v_1,v_2}(m_2)(v_2 - v_1 m_2)$, where $c_{v_1,v_2}(m_2) > 0$ for any m_2, and that

$$f(\pm\infty) = \lim_{m_2\to\pm\infty} f(m_2) = \frac{\sqrt{1-v_1^2} \pm v_2}{\sqrt{1-v_1^2} \mp v_2}.$$

We treat the cases $v_1 = 0$ and $v_1 \neq 0$ separately.

(a) Assume that $v_1 = 0$. If $v_2 = 0$, then $V = I_2$ and Theorem 2.5 establishes the result. If $v_2 \neq 0$, then f has no critical point and

$$\sup_{m_2\in\mathbb{R}} f(m_2) = \max\big(f(-\infty), f(\infty)\big) = \frac{1+|v_2|}{1-|v_2|} \text{ and } \inf_{m_2\in\mathbb{R}} f(m_2) = \min\big(f(-\infty), f(\infty)\big) = \frac{1-|v_2|}{1+|v_2|},$$

so that (B.12) yields

$$D_0(V,P) = P\left[\frac{X_1^2}{X_2^2} \geq \frac{1+|v_2|}{1-|v_2|}\right] = P\left[\frac{X_1^2}{X_2^2} \geq \frac{1+(1-\det V)^{1/2}}{1-(1-\det V)^{1/2}}\right] = P\left[Y_2 \geq \frac{1}{2} + \frac{1}{2}\{1-\det(V)\}^{1/2}\right],$$

where we have used the fact that if Z has a $F(1,1)$ Fisher-Snedecor distribution, then $Z/(1+Z)$ has a Beta$(1/2, 1/2)$ distribution.

(b) Assume now that $v_1 \neq 0$. Then the only critical point of f is $m_2^{\mathrm{crit}} = v_2/v_1$, so that, irrespective of the fact that this critical point is a local minimum/maximum of f,

$$\sup_{m_2\in\mathbb{R}} f(m_2) = \max\big(f(-\infty), f(\infty), f(m_2^{\mathrm{crit}})\big) = \max\left(\frac{(1-v_1^2)^{1/2} + |v_2|}{(1-v_1^2)^{1/2} - |v_2|}, \frac{\mathrm{Sign}(v_1) + (v_1^2+v_2^2)^{1/2}}{\mathrm{Sign}(v_1) - (v_1^2+v_2^2)^{1/2}}\right)$$

and

$$\inf_{m_2\in\mathbb{R}} f(m_2) = \min\big(f(-\infty), f(\infty), f(m_2^{\mathrm{crit}})\big) = \min\left(\frac{(1-v_1^2)^{1/2} - |v_2|}{(1-v_1^2)^{1/2} + |v_2|}, \frac{\mathrm{Sign}(v_1) + (v_1^2+v_2^2)^{1/2}}{\mathrm{Sign}(v_1) - (v_1^2+v_2^2)^{1/2}}\right).$$

Lemma B.6 yields

$$\sup_{m_2\in\mathbb{R}} f(m_2) = \frac{(1-v_1^2)^{1/2} + |v_2|}{(1-v_1^2)^{1/2} - |v_2|}\mathbb{I}[v_1 < 0] + \frac{1 + (v_1^2+v_2^2)^{1/2}}{1 - (v_1^2+v_2^2)^{1/2}}\mathbb{I}[v_1 > 0]$$

and

$$\inf_{m_2\in\mathbb{R}} f(m_2) = \frac{-1+(v_1^2+v_2^2)^{1/2}}{-1-(v_1^2+v_2^2)^{1/2}}\,\mathbb{I}[v_1<0] + \frac{(1-v_1^2)^{1/2}-|v_2|}{(1-v_1^2)^{1/2}+|v_2|}\,\mathbb{I}[v_1>0],$$

hence also

$$\max\left(\sup_{m_2\in\mathbb{R}} f(m_2)\ ,\ 1/\inf_{m_2\in\mathbb{R}} f(m_2)\right) = \frac{1+(v_1^2+v_2^2)^{1/2}}{1-(v_1^2+v_2^2)^{1/2}} = \frac{1+(1-\det(V))^{1/2}}{1-(1-\det(V))^{1/2}}.$$

Therefore, (B.12) finally provides

$$D_0(V,P) = P\left[\frac{X_1^2}{X_2^2} \geq \frac{1+(1-\det V)^{1/2}}{1-(1-\det V)^{1/2}}\right] = P\left[Y_2 \geq \frac{1}{2}+\frac{1}{2}\big\{1-\det(V)\big\}^{1/2}x\right].$$

This proves the result for the case where P is bivariate standard normal. The general result then follows from Part (i) of the Theorem. □

B.2 Proofs from Sections 3 and 5

Proof of Theorem 3.1. Let P and Q be two probability measures over $\mathbb{R}^k$ and fix $V \in \mathcal{P}_{k,\mathrm{tr}}$. Fix $\varepsilon > 0$ and assume, without loss of generality, that $D_\theta(V,P) \leq D_\theta(V,Q)$. Lemma B.1 entails that there exists $M_0 \in \mathcal{M}_k^0$ such that $P\big[C_{\theta,V}^{M_0}\big] \leq D_\theta(V,P)+\varepsilon$, where we still use the notation $C_{\theta,V}^M = \big\{x \in \mathbb{R}^k \setminus \{\theta\} : (u_{\theta,V}^x)'Mu_{\theta,V}^x \geq \frac{1}{k}\mathrm{tr}(M)\big\}$. Consequently, using Lemma B.1 again,

$$\begin{aligned}|D_\theta(V,Q)-D_\theta(V,P)| &= D_\theta(V,Q)-D_\theta(V,P)\\ &\leq Q\big[C_{\theta,V}^{M_0}\big] - P\big[C_{\theta,V}^{M_0}\big] + \varepsilon \leq \sup_{C\in\mathcal{C}_\theta} |Q[C]-P[C]| + \varepsilon,\end{aligned}$$

with $\mathcal{C}_\theta = \{C_{\theta,V}^M : M \in \mathcal{M}_k^0, V \in \mathcal{P}_{k,\mathrm{tr}}\}$. Since this holds for any $\varepsilon > 0$ and for any $V \in \mathcal{P}_{k,\mathrm{tr}}$, we have

$$\sup_{V\in\mathcal{P}_{k,\mathrm{tr}}} |D_\theta(V,Q)-D_\theta(V,P)| \leq \sup_{C\in\mathcal{C}_\theta} |Q[C]-P[C]|.$$

It thus only remains to show that $\mathcal{C}_\theta$ is a Vapnik-Chervonenkis class. To do so, note that $C_{\theta,V}^M = \big\{x \in \mathbb{R}^k \setminus \{\theta\} : (x-\theta)'V^{-1/2}MV^{-1/2}(x-\theta) \geq 0\big\}$, so that $\mathcal{C}_\theta \subset \{D_{\theta,A} \cap (\mathbb{R}^k \setminus \{\theta\}) : A \in \mathcal{M}_k^{\mathrm{all}}\}$, with $D_{\theta,A} = \big\{x \in \mathbb{R}^k : (x-\theta)'A(x-\theta) \geq 0\big\}$. Theorem 4.6 from Dudley (2014) implies that $\{D_{\theta,A} : A \in \mathcal{M}_k^{\mathrm{all}}\}$ is a Vapnik-Chervonenkis class $\mathcal{D}_\theta$. It then follows from Lemma 2.6.17(ii) in Van der Vaart & Wellner (1996) that $\{D_{\theta,A} \cap (\mathbb{R}^k \setminus \{\theta\}) : A \in \mathcal{M}_k^{\mathrm{all}}\}$, hence also $\mathcal{C}_\theta$, is a Vapnik-Chervonenkis class. □

Proof of Theorem 3.2. Recall from (2.1) that $V_{\theta,P}$ is defined as the barycentre of $R_\theta(\alpha_*,P)$, with $\alpha_* = \max_V D_\theta(V,P)$. The mapping $V \mapsto D_\theta(V,P)$ is upper semicontinuous (Theorem 2.1) and constant over $R_\theta(\alpha_*,P)$. Clearly, it is easy to define a mapping $V \mapsto \tilde{D}_\theta(V,P)$ that is upper semicontinuous, agrees with $V \mapsto D_\theta(V,P)$ in the complement of $R_\theta(\alpha_*,P)$, and for which $V_{\theta,P}$ is the unique maximiser. By using Theorem 3.1, the result then follows from Theorem 2.12 and Lemma 14.3 in Kosorok (2008). □

The proof of Theorem 6.1(i) is long and technical, but follows along the same lines as the proof of Theorem 2.2 from Chapter I. As for the proof of Theorem 6.1(ii), it is strictly the same as that of Theorem 3.2. We therefore omit the proof of Theorem 6.1.

We finish with proving the non Fisher consistency of Mizera's tangent depth, as explained in Section 5.

Lemma B.7. Let P be the distribution of a bivariate normal random vector with mean $\theta_0 = 0_2$, scale $\sigma_0 = tr(\Sigma)/2 = 1$, Gaussian radial density $g = \phi$ and shape $V_{3/4}$, where $V_a = \mathrm{diag}(a, 2-a)$. Then $D(V_{3/4}, P) < D(V_1, P)$, where I_2 denotes the 2×2 identity matrix.

Proof of Lemma B.7. Parallel to the proof of Theorem B.1 above, it can be showed that, denoting $W_a = \Big(2X_1X_2, \frac{X_2^2}{(2-a)^2} - \frac{X_1^2}{a^2} - \big(\frac{1}{2-a} - \frac{1}{a}\big)\Big)'$ and using $\varphi_g(z) = z$ and $\sigma_S = 1$, it holds $D(V_a, P) = HD(0_2, P_{W_a})$.

We start by computing the depth of $V_1 = I_2$, on the basis of W_1. The spherical symmetry of the random vector $Z = (Z_1, Z_2)' = (X_1/\sqrt{a_0}, X_2/\sqrt{2-a_0})'$ yields

$$\begin{aligned} D(V_1, P) &= HD(0_2, P_{W_1}) = \inf_{w\in\mathbb{R}^2} P[w'W_1 \geq 0] \\ &= \inf_{w\in\mathbb{R}^2} P[Z'Q_wZ \geq 0] = \inf_{w\in\mathbb{R}^2} P[\lambda_+^w Z_1^2 + \lambda_-^w Z_2^2 \geq 0], \end{aligned} \tag{B.13}$$

where $\lambda_\pm^w = (1-a_0)w_2 \pm \sqrt{(2a_0 - a_0^2)w_1^2 + w_2^2}$ are the eigenvalues of

$$Q_w := \begin{pmatrix} -a_0 w_2 & \sqrt{a_0(2-a_0)}w_1 \\ \sqrt{a_0(2-a_0)}w_1 & (2-a_0)w_2 \end{pmatrix}.$$

The last infimum in (B.13) can be taken over all w vectors of the form $w = w(\theta) = (\cos(\theta)/\sqrt{2a_0 - a_0^2}, \sin(\theta))'$, for $\theta \in [0, 2\pi]$. For these $w(\theta)$, one has $\lambda_\pm^w = (1-a_0)\sin(\theta) \pm 1$, and the infimum is obtained for $\sin\theta = -1$, which corresponds to $w = (0, -1)'$. The depth of $V_1 = I_2$ is therefore given by

$$D(V_1, P) = P\big[a_0 Z_1^2 + (a_0 - 2)Z_2^2 \geq 0\big] =: c_{a_0}(1). \tag{B.14}$$

Turning to the depth of V_{a_0}, we of course have that

$$D(V_{a_0}, P) \leq P[(W_{a_0})_2 \leq 0] = P\big[a_0 Z_2^2 + (a_0 - 2)Z_1^2 \leq (2a_0 - 2)\big] =: c_{a_0}(a_0),$$

where $Z = (Z_1, Z_2)'$ is as above. By conditioning on Z_2^2, one can show that

$$c_{a_0}(1) = 1 - \int_0^\infty F\Big(\frac{(2-a_0)z}{a_0}\Big) f(z)\, dz$$

and

$$c_{a_0}(a_0) = \int_{\frac{2-2a_0}{2-a_0}}^\infty F\Big(\frac{(2a_0-2) + (2-a_0)z}{a_0}\Big) f(z)\, dz,$$

where $F(\,\cdot\,)$ and $f(\,\cdot\,)$ are the cdf and pdf of the χ_1^2 distribution, respectively. For $a_0 = .75$, this yields that $D(V_{a_0}, P) \leq c_{a_0}(a_0) \approx .3732$ is indeed strictly smaller than $D(V_1, P) = c_{a_0}(1) \approx .4196$. □

We finish this section by proving the following result on the dimensionality reduction available in computing Tyler shape depth.

Theorem B.1. Let $P = P^X$ be a probability measure over $\mathbb{R}^k$ and fix $\theta \in \mathbb{R}^k$. Then, $D_\theta(V, P) = D(0, P^{\tilde{W}_{\theta_P, V}}) = \inf_{u \in \mathcal{S}^{d_k - 1}} P\big[u'\tilde{W}_{\theta,V} \geq 0\big]$, with $\tilde{W}_{\theta,V} = \mathrm{vech}_0\big(U_{\theta,V}U'_{\theta,V} - \frac{1}{k}I_k\big)$.

Proof of Theorem B.1. Write $L_{\theta,V} = U_{\theta,V}U'_{\theta,V} - \frac{1}{k}I_k$. Since $(L_{\theta,V})_{11} = -\sum_{\ell=2}^k (L_{\theta,V})_{\ell\ell}$, there exists a $(d_k + 1) \times d_k$ full-rank matrix H_0 such that $\mathrm{vech}(L_{\theta,V}) = H_0 \,\mathrm{vech}_0(L_{\theta,V})$. Therefore, there exists a $k^2 \times d_k$ full-rank matrix H (one can take $H = DH_0$, where D is the usual duplication matrix) such that $W_{\theta,V} = \mathrm{vec}(L_{\theta,V}) = H\tilde{W}_{\theta,V}$. It follows that

$$D_\theta(V, P) = D(0, P^{W_{\theta,V}}) = \inf_{u \in \mathbb{R}^{k^2}} P\big[u'W_{\theta,V} \geq 0\big]$$

$$= \inf_{u \in \mathbb{R}^{k^2}} P\big[(H'u)'\,\tilde{W}_{\theta,V} \geq 0\big] = \inf_{v \in \mathbb{R}^{d_k}} P\big[v'\tilde{W}_{\theta,V} \geq 0\big] = D(0, P^{\tilde{W}_{\theta,V}}),$$

where we used the fact that H' has full column rank. □

C Acknowledgments

The author's research is supported by the FNRS FC8444 "Chargé de Recherches" grant and their support is gratefully acknowledged. The author also wishes to thank Professor Davy Paindaveine for the collaboration and enriching discussions on both chapters of this thesis.

General Bibliography

Arcones, M. A. & Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.*, *21*(3), 1494–1542.

Barnett, V. (1976). The ordering of multivariate data. *J. Roy. Statist. Soc. Ser. A*, *139*(3), 318–355. With a discussion by R. L. Plackett, K. V. Mardia, R. M. Loynes, A. Huitson, G. M. Paddle, T. Lewis, G. A. Barnard, A. M. Walker, F. Downton, P. J. Green, Maurice Kendall, A. Robinson, Allan Seheult and D. H. Young.

Bartoszyński, R., Pearl, D., & Lawrence, J. (1997). A multidimensional goodness-of-fit test based on interpoint distances. *J. Amer. Statist. Assoc.*, *92*, 577–586.

Berger, M. (2003). *A Panoramic View of Riemannian Geometry*. Berlin Heidelberg: Springer-Verlag.

Bhatia, R. (2007). *Positive definite matrices*. Princeton Series in Applied Mathematics. Princeton, NJ: Princeton University Press.

Bhatia, R. & Holbrook, J. (2006). Riemannian geometry and matrix geometric means. *Linear Algebra Appl.*, *413*, 594–618.

Bickel, P. J. (1964). On some alternative estimates for shift in the p-variate one sample problem. *Ann. Math. Statist.*, *35*, 1079–1090.

Brown, B. M. (1983). Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B*, *45*(1), 25–30.

Cardot, H., Cénac, P., & Godichon-Baggioni, A. (2017). Online estimation of the geometric median in hilbert spaces : Non asymptotic confidence balls. *Ann. Statist.*, *45*, 591–614.

Carrizosa, E. (1996). A characterization of halfspace depth. *J. Multivariate Anal.*, *58*, 21–26.

Cartan, E. (1929). Groupes simples clos et ouverts et géometrie riemannienne. *J. Math. Pures Appl.*, *8*, 1–33.

Cascos, I. (2009). Data depth: Multivariate statistics and geometry. In W. Kendall & I. Molchanov (Eds.), *New perspectives in stochastic geometry*. Clarendon Press, Oxford University Press, Oxford.

Chakraborty, A. & Chaudhuri, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann. Statist.*, *42*, 1203–1231.

Chakraborty, B. & Chaudhuri, P. (1996). On a transformation and re-transformation technique for constructing an affine equivariant multivariate median. *Proc. Amer. Math. Soc.*, *124*(8), 2539–2547. URL `http://dx.doi.org/10.1090/S0002-9939-96-03657-X`.

Chakraborty, B. & Chaudhuri, P. (1998). On an adaptive transformation-retransformation estimate of multivariate location. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, *60*(1), 145–157. URL `http://dx.doi.org/10.1111/1467-9868.00114`.

Chakraborty, B., Chaudhuri, P., & Oja, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statist. Sinica*, *8*(3), 767–784.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.*, *91*(434), 862–872.

Chen, M., Gao, C., & Ren, Z. (2017). Robust covariance and scatter matrix estimation under huber's contamination model. *Ann. Statist.*, to appear.

Claeskens, G., Hubert, M., Slaets, L., & Vakili, K. (2014). Multivariate functional halfspace depth. *J. Amer. Statist. Assoc.*, *109*, 411–423.

Cuevas, A., Febrero, M., & Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.*, *22*, 481–496.

Dang, X. & Serfling, R. J. (2010). Nonparametric depth-based multivariate outlier identifiers, and masking robustness properties. *J. Statist. Plann. Inference*, *140*, 198–213.

Donoho, D. (1982). *Breakdown properties of multivariate location estimators*. Ph.D. thesis, Department of Statistics, Harvard University.

Donoho, D. & Gasko, M. (1987). *Multivariate Generalizations of the Median and Trimmed Mean, I*. Technical report (University of California, Berkeley. Dept. of Statistics). Department of Statistics, University of California. URL `http://books.google.be/books?id=LzS3twAACAAJ`.

Donoho, D. L. & Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.*, *20*(4), 1803–1827.

Dudley, R. M. (2014). *Uniform Central Limit Theorems*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2nd ed.

Dudley, R. M., Sidenko, S., & Wang, Z. (2009). Differentiability of t-functionals of location and scatter. *Ann. Statist.*, *37*, 939–960.

Dümbgen, L. (1998). On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.*, *50*(3), 471–491.

Dümbgen, L., Pauly, M., & Schweizer, T. (2015). M-functionals of multivariate scatter. *Statist. Surv.*, *9*, 32–105.

Dümbgen, L. & Tyler, D. E. (2016). Geodesic convexity and regularized scatter estimators, ArXiv:1607.05455v2.

Dürer, A. (1525). *A course in the art of measurement with compass and ruler: Underweysung der Messung*. Alan Wofsy Fine Arts, 1981.

Dyckerhoff, R. (2002). Inference based on data depth. In *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach*, Springer, New York. 133–163.

Elmore, R. T., Hettmansperger, T. P., & Xuan, F. (2006). Spherical data depth and a multivariate median. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, Amer. Math. Soc., vol. 72 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*. 87–101.

Fan, Y., Jin, J., & Yao, Z. (2013). Optimal classification in sparse gaussian graphical model. *Ann. Statist.*, *41*, 2537–2571.

Fan, Y. & Lv, J. (2016). Innovated scalable efficient estimation in ultra-large gaussian graphical models. *Ann. Statist.*, *44*, 2098–2126.

Fraiman, R., Liu, R. Y., & Meloche, J. (1997). Multivariate density estimation by probing depth. In *L_1-statistical procedures and related topics (Neuchâtel, 1997)*, Hayward, CA: Inst. Math. Statist., vol. 31 of *IMS Lecture Notes Monogr. Ser.*. 415–430. URL http://dx.doi.org/10.1214/lnms/1215454155.

Fraiman, R. & Meloche, J. (1999). Multivariate l-estimation (with discussion). *Test*, *8*, 255–317.

Gao, Y. (2003). Data depth based on spatial ranks. *Statist. Probab. Lett.*, *65*, 217–225.

Gauss, C. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*, Hamburg: Perthes et Besser. English translation by C.H. Davis, reprinted by Dover, New York (1963).

Gini, C. & Galvani, L. (1929). Di talune estensioni, dei concetti di media ai caratteri qualitativi. *Metron*, *8*, 3–209.

Gower, J. (1974). Algorithm as78. the mediancenter. *Applied Statistics*, *23*, 466–470.

Green, P. (1981). Peeling bivariate data. In V. Barnett (Ed.), *Interpreting Multivariate Data*, Wiley, New York. 3–20.

Haldane, J. (1948). Note on the median of a multivariate distribution. *Biometrika*, *35*(3/4), 414–415.

Hall, P. & Jin, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.*, *38*, 1686–1732.

Hallin, M. & Paindaveine, D. (2006a). Parametric and semiparametric inference for shape: the role of the scale functional. *Statist. Decisions*, *24*(3), 327–350.

Hallin, M. & Paindaveine, D. (2006b). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.*, *34*(6), 2707–2756.

Hallin, M., Paindaveine, D., & Siman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.*, *38*, 635–669.

Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.*, *42*, 1887–1896.

Hardy, G. H. (1940). *A mathematician's apology*. Cambridge University Press.

Hassairi, A. & Regaieg, O. (2008). On the tukey depth of a continuous probability distribution. *Statist. Probab. Lett.*, *78*, 2308–2313.

Hayford, J. (1902). What is the center of an area, or the center of a population? *J. Amer. Statist. Assoc.*, *8*, 47–58.

He, X., Simpson, D., & Portnoy, S. (1990). Breakdown robustness of tests. *J. Amer. Statist. Assoc.*, *85*, 446–452.

He, Y. & Einmahl, J. (2017). Estimation of extreme depth-based quantile regions. *J. R. Stat. Soc. Ser. B*, *79*, 449–461.

Hettmansperger, T. P., Nyblom, J., & Oja, H. (1992). On multivariate notions of sign and rank. In Y. Dodge (Ed.), L_1 *Statistical and Related Methods*. North-Holland, Amsterdam, 267–278.

Hettmansperger, T. P. & Randles, R. H. (2002). A practical affine equivariant multivariate median. *Biometrika*, *89*(4), 851–860.

Hodges, J. (1955). A bivariate sign test. *Ann. Math. Statist.*, *26*, 523–527.

Hodges, J. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. In *Proc. 5th Berkeley Symp.*. vol. 1, 163–178.

Hojsgaard, S., Edwards, D., & Lauritzen, S. (2012). *Graphical models with R*. Springer-Verlag New York.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, *39*, 41–57.

Hu, Y., Wang, Y., & Wu, Y. (2011). Tensor-based projection depth. *Bernoulli*, *17*(4), 1386–1399. URL `http://dx.doi.org/10.3150/10-BEJ317`.

Hubert, M., Rousseeuw, P. J., & Segaert, P. (2015). Multivariate functional outlier detection. *Stat. Methods Appl.*, *24*, 177–202.

Hugg, J., Rafalin, E., Seyboth, K., & Souvaine, D. (2006). An experimental study of old and new depth measures. In *In Proc. Workshop on Algorithm Engineering and Experiments , Lecture Notes in Computer Science*. Springer, 51–64.

Ilmonen, P. & Paindaveine, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Ann. Statist.*, *39*, 2448–2476.

John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika*, *59*, 169–173.

Kemperman, J. H. B. (1987). The median of a finite measure on a Banach space. In *Statistical data analysis based on the* L_1*-norm and related methods (Neuchâtel, 1987)*, Amsterdam: North-Holland. 217–230.

Kent, J. & Tyler, D. E. (1988). Maximum likelihood estimation for the wrapped cauchy distribution. *J. Appl. Statist.*, *15*, 247–254.

Kent, J. & Tyler, D. E. (1991). Redescending M-estimates of multivariate location and scatter. *Ann. Statist.*, *19*, 2102–2119.

Kong, L. & Zuo, Y. (2010). Smooth depth contours characterize the underlying distribution. *J. Multivariate Anal.*, *101*(9), 2222–2226. URL `http://dx.doi.org/10.1016/j.jmva.2010.06.007`.

Koshevoy, G. & Mosler, K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.*, *25*(5), 1998–2017. URL http://dx.doi.org/10.1214/aos/1069362382.

Koshevoy, G. A. (2002). The Tukey depth characterizes the atomic measure. *J. Multivariate Anal.*, *83*(2), 360–364. URL http://dx.doi.org/10.1006/jmva.2001.2052.

Koshevoy, G. A. (2003). Lift-zonoid and multivariate depths. In *Developments in Robust Statistics. International Conference on Robust Statistics 2001 (ICORS)*.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. New York: Springer.

Kuhn, H. (1973). A note on Fermat's problem. *Math. Program.*, *4*, 98–107.

Lawson, J. & Lim, Y. (2013). Weighted means and karcher equations of positive operators. *Proc. Natl. Acad. Sci. USA*, *110*, 15626–15632.

Liu, R. Y. (1987). Simplicial depth and the related location estimators. Tech. rep., Dept. Statist. Rutgers Univ.

Liu, R. Y. (1988). On a notion of simplicial depth. *Proc. Nat. Acad. Sci. U.S.A.*, *85*(6), 1732–1734. URL http://dx.doi.org/10.1073/pnas.85.6.1732.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.*, *18*(1), 405–414.

Liu, R. Y. (1992). Data depth and multivariate rank tests. In Y. Dodge (Ed.), L_1 *Statistics and Related Methods*. North-Holland, Amsterdam, 279–294.

Liu, R. Y., Parelius, J. M., & Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.*, *27*(3), 783–858.

Liu, R. Y., Serfling, R. J., & Souvaine, D. L. (Eds.) (2006). *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*. Amer. Math. Soc.

Liu, R. Y. & Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.*, *88*(421), 252–260. URL http://links.jstor.org/sici?sici=0162-1459(199303)88:421<252:AQIBOD>2.0.CO;2-Y&origin=MSN.

Liu, Z. & Modarres, R. (2011). Lens data depth and median. *J. Nonparametr. Stat.*, *23*, 1063–1074.

Lok, W. & Lee, S. M. S. (2011). A new statistical depth function with applications to multimodal data. *J. Nonparametr. Stat.*, *23*(3), 617–631.

López-Pintado, S. & Romo, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.*, *104*(486), 718–734.

Milasevic, P. & Ducharme, G. R. (1987). Uniqueness of the spatial median. *Ann. Statist.*, *15*(3), 1332–1333. URL http://dx.doi.org/10.1214/aos/1176350511.

Mizera, I. (2002). On depth and deep points: a calculus. *Ann. Statist.*, *30*(6), 1681–1736.

Mizera, I. & Müller, C. H. (2004). Location-scale depth (with discussion). *J. Amer. Statist. Assoc.*, *99*(468), 949–989.

Mizera, I. & Volauf, M. (2002). Continuity of halfspace depth contours and maximum depth estimators: diagnostics of depth-related methods. *J. Multivariate Anal.*, *83*(2), 365–388. URL `http://dx.doi.org/10.1006/jmva.2001.2053`.

Mosler, K. (2012). Data depth. Arxiv:1207-4988.

Mosler, K. & Hoberg, R. (2006). Data analysis and classification with the zonoid depth. In *Data depth: robust multivariate analysis, computational geometry and applications*, Providence, RI: Amer. Math. Soc., vol. 72 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*. 49–59.

Mosteller, C. & Tukey, J. W. (1977). *Data analysis and regression*. Addison-Wesley.

Müller, C. H. (2005). Depth estimators and tests based on the likelihood principle with application to regression. *J. Multivariate Anal.*, *95*(1), 153–181.

Nieto-Reyes, A. & Battey, H. (2016). A topologically valid definition of depth for functional data. *Statist. Sci.*, *31*, 61–79.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.*, *1*(6), 327–332. URL `http://dx.doi.org/10.1016/0167-7152(83)90054-8`.

Oja, H., Sirkiä, S., & Eriksson, J. (2006). Scatter matrices and independent component analysis. *Australian Journal of Statistics*, *35*, 175–189.

Paindaveine, D. (2008). A canonical definition of shape. *Statist. Probab. Lett.*, *78*(14), 2240–2247.

Paindaveine, D. & Van Bever, G. (2014). Inference on the shape of elliptical distributions based on the mcd. *J. Multivariate Anal.*, *129*, 125–144.

Paindaveine, D. & Van Bever, G. (2015). Nonparametrically consistent depth-based classifiers. *Bernoulli*, *21*, 62–82.

Paindaveine, D. & Van Bever, G. (2017a). Halfspace depths for scatter, concentration and shape matrices. *Ann. Statist.*, to appear.

Paindaveine, D. & Van Bever, G. (2017b). On the maximal halfspace depth of permutation-invariant distributions on the simplex. *Statist. Probab. Lett.*, *129*, 335–339.

Plackett, R. (1976). Comment on "the ordering of multivariate data" by v. barnett. *J. Roy. Statist. Soc. Ser. A*, *139*, 344–346.

Romanazzi, M. (2009). Data depth, random simplices and multivariate dispersion. *Statist. Probab. Lett.*, *79*(12), 1473–1479. URL `http://dx.doi.org/10.1016/j.spl.2009.03.022`.

Rousseeuw, P. J. & Hubert, M. (1999). Regression depth (with discussion). *J. Amer. Statist. Assoc.*, *94*(446), 388–433.

Rousseeuw, P. J. & Ruts, I. (1999). The depth function of a population distribution. *Metrika*, *49*(3), 213–244.

Rousseeuw, P. J. & Struyf, A. (2004). Characterizing angular symmetry and regression symmetry. *J. Statist. Plann. Inference*, *122*(1-2), 161–173.

Scates, D. (1933). Locating the median of the population in the united states. *Metron*, *11*, 49–66.

Serfling, R. (2004a). Nonparametric multivariate descriptive measures based on spatial quantiles. *J. Statist. Plann. Inference*, *123*(2), 259–278. URL `http://dx.doi.org/10.1016/S0378-3758(03)00156-3`.

Serfling, R. (2006). Depth functions in nonparametric multivariate inference. In *Data depth: robust multivariate analysis, computational geometry and applications*, Providence, RI: Amer. Math. Soc., vol. 72 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*. 1–16.

Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization. *J. Nonparametr. Stat.*, *22*, 915–926.

Serfling, R. J. (2002). Quantile functions for multivariate analysis: Approaches and applications. *Stat. Neerl.*, *56*(2), 214–232. Special issue: Frontier research in theoretical statistics, 2000 (Eindhoven).

Serfling, R. J. (2004b). Some perspectives on location and scale depth functions. *J. Amer. Statist. Assoc.*, *99*, 970–973.

Simpson, R. (1750). *Doctrine and Application of Fluxions*. Printed for John Nourse, London.

Small, C. G. (1987). Measures of centrality for multivariate and directional distributions. *Canad. J. Statist.*, *15*, 31–39.

Small, C. G. (1990). A survey of multidimensional medians. *Internat. Statist. Rev.*, *58*(3), 263–277.

Stahel, W. A. (1981). *Robuste Schaätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen*. Ph.D. thesis, Eidgenossiche Technische Hochschule, Zürich.

Struyf, A. & Rousseeuw, P. J. (1999). Halfspace depth and regression depth characterizes the empirical distribution. *J. Multivariate Anal.*, *69*, 135–153.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*. Canad. Math. Congress, Montreal, Que., 523–531.

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesly.

Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.*, *15*(1), 234–251.

Tyler, D. E. (1994). Finite-sample breakdown points of projection based multivariate location and scatter statistics. *Ann. Statist.*, *22*, 1024–1044.

Van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. New York: Springer.

Vardi, Y. & Zhang, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proc. Natl. Acad. Sci. USA*, *97*(4), 1423–1426.

Vardi, Y. & Zhang, C.-H. (2001). A modified Weiszfeld algorithm for the Fermat-Weber location problem. *Math. Program., Ser. A*, *90*, 559–566.

Weber, A. (1909). Über den standort der industrien, tubingen. In *Alfred Weber's theory of location of industries*, University of Chicago Press. English translation by Friedriech C.J. (1929).

Zhang, J. (2002). Some extensions of Tukey's depth function. *J. Multivariate Anal.*, *82*(1), 134–165.

Zuo, Y. (2003a). Projection-based depth functions and associated medians. *Ann. Statist.*, *31*(5), 1460–1490.

Zuo, Y. (2003b). Projection-based depth functions and associated medians. *Ann. Statist.*, *31*(5), 1460–1490.

Zuo, Y. & Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.*, *28*(2), 461–482.